Durham University

Durham Research Online

Deposited in DRO:

17 October 2018

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Kneuer, Carsten and Charistou, Agathi and Craig, Peter and Eleftheriadou, Dimitra and Engel, Nadine and Kjaerstad, Mia and Krishnan, Shaji and Laskari, Vasileia and Machera, Kyriaki and Nikolopoulou, Dimitra and Pieper, Christina and Schoen, Eric and Spilioti, Eliana and Buist, Harrie (2018) 'Applicability of in silico tools for the prediction of dermal absorption for pesticides.', EFSA supporting publications., 15 (10). 1493E.

Further information on publisher's website:

https://doi.org/10.2903/sp.efsa.2018.EN-1493

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.

Durham University Library, Stockton Road, Durham DH1 3LY, United Kingdom
Tel : +44 (0)191 334 3042 | Fax : +44 (0)191 334 2971
https://dro.dur.ac.uk

EXTERNAL SCIENTIFIC REPORT

APPROVED: 01 October 2018

doi:10.2903/sp.efsa.2018.EN-1493

Applicability of *in silico* tools for the prediction of dermal absorption for pesticides

Carsten Kneuer, Agathi Charistou, Peter Craig, Dimitra Eleftheriadou, Nadine Engel, Mia Kjaerstad, Shaji Krishnan, Vasileia Laskari, Kyriaki Machera, Dimitra Nikolopoulou, Christina Pieper, Eric Schoen, Eliana Spilioti and Harrie Buist

German Federal Institute for Risk Assessment (BfR), Berlin, Germany; BENAKI Phytopathological Institute (BPI), Kifissia, Athens, Greece; Danish EPA, Copenhagen, Denmark; University of Durham, Durham, UK and TNO innovation for life (TNO), Zeist, The Netherlands

Abstract

Based on the "Human *in vitro* dermal absorption datasets" published as supporting information to the revised EFSA Guidance on Dermal Absorption, *in silico* models for prediction of absorption across the skin have been evaluated. For this evaluation, a systematic literature search and review was performed, identifying 288 publications describing mathematical models for prediction of dermal absorption. Eleven models potentially relevant to the regulatory assessment of pesticides and which cover a range of approaches were selected for in depth evaluation. This included three mixture models taking into account physicochemical properties of the co-formulants such as polar surface area, hydrogen bonding or octanol-water partition coefficients. Additional data on the pesticidal active substances and information on the composition of some of the formulations covered in the dermal absorption dataset were gathered, as these were required as input parameters for the selected models. The models were implemented with settings reflecting as much as possible realistic exposure scenarios and the experimental conditions under which the measured data were obtained. As the majority of the models predicted either maximum flux or the permeation coefficient, further combination with a model achieving translation into percentage absorption was required. This was done with and without consideration of the lag time. Only one model directly predicted percentage dermal absorption, which was a spreadsheet-based single substance model taking into consideration several skin parameters, experimental conditions, various physicochemical properties of the active substance and the type of vehicle. Statistical analysis of model predictions revealed overall low concordance with measured values, thereby limiting regulatory acceptance. Additional analysis was performed on the results of two mixture models and the above mentioned complex single substance model which showed moderate correlation between predicted and measured data. Options to improve model performance were discussed and Bayesian random effects modelling was explored to adjust predicted percentage dermal absorption to measured data as was model combination. When taking into account observed uncertainties of predictions, one of the models may provide a Tier 2 tool to estimate dermal absorption value in the absence of adequate experimental data when the predicted values are in the range of 10 to 70%. However, further work is needed to better understand the effect of co-formulants on dermal absorption of pesticides and to improve model predicitivity.

© European Food Safety Authority, 2018

Key words: dermal absorption, *in silico*, formulations, mixtures, pesticides, (Q)SAR, skin penetration

Question number: EFSA-Q-2016-00516

Correspondence: pras.secretariat@efsa.europa.eu

Disclaimer: The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Acknowledgements: The authors wish to acknowledge the respondents to the survey which has instructed this analysis. The authors also wish to thank EFSA staff members Arianna Chiusolo and Rositsa Serafimova for their technical support and EFSA for financing this project.

Suggested citation: Carsten Kneuer, Agathi Charistou, Peter Craig, Dimitra Eleftheriadou, Nadine Engel, Mia Kjaerstad, Shaji Krishnan, Vasileia Laskari, Kyriaki Machera, Dimitra Nikolopoulou, Christina Pieper, Eric Schoen, Eliana Spilioti and Harrie Buist, 2018. Applicability of in silico tools for the prediction of dermal absorption for pesticides. EFSA supporting publication 2018:EN-1493. 156 pp. doi:10.2903/sp.efsa.2018.EN-1493

ISSN: 2397-8325

© European Food Safety Authority, 2018

Summary

A search protocol to retrieve papers on models predicting dermal absorption from public literature for application in Scopus, PubMed and ToxLine was developed in cooperation with EFSA. In addition, an internet search using Google was performed. The records retrieved were unduplicated and transferred into the TNO Literature Review Tool (LRT) for relevance check. The 2212 records collected yielded 288 papers considered relevant based on their title and abstract and copies of these papers were obtained for further scrutiny.

Subsequently, a first review of the scientific quality and of the relevance for the regulatory framework of pesticide risk assessment of the models described in the selected papers was executed according to a protocol refined in cooperation with EFSA. Issues considered were the descriptors and algorithms used, the provision of details on the training sets employed for model development and of actual input values, the statistical parameters provided to illustrate the goodness-of-fit, robustness and predictivity of the models and the extent to which the models offered a mechanistic interpretation of the dermal absorption process.

Based on the first stage of the scientific review, 103 out of the 288 papers originally regarded as relevant were excluded from further consideration as they proved not to be relevant when the full publication was assessed. Only the publications describing original work were entered for the second stage of the scientific review, as the review papers were not expected to contribute additions to the pool of models extracted from the original studies. The 142 original papers described a total of 233 models, 188 predicting absorption of single substances and 45 predicting absorption of substances from mixtures. Most models were algorithm based: 141 out of 188 single substance models and 37 out of 45 mixture models; the remainder were machine learning models.

In the second stage of the scientific review, models were screened for their potential usefulnes regarding the prediction of dermal absorption of pesticides and their dilutions in water or of pesticides from mixtures (including solutions in vehicles other than water). Of the 111 models identified as possibly useful, 95 models predict the permeation contstant (k_p), 15 provide estimates for either maximum or steady state flux (J_{max} and J_{SS}, respectively), whereas only one predicted percentage absorption.

A selection of models for prediction of dermal absorption was made for comparison of predicted versus measured data provided in the "EFSA Human *in vitro* dermal absorption dataset". A selection proposal was presented to and agreed with EFSA. In accordance with the specifications, eleven models including at three models taking into account physicochemical properties of the co-formulants such as polar surface area, hydrogen bonding or octanol-water partition coefficients (hereafter called "mixture models") were selected for implementation and statistical evaluation. Models predicting either flux (J_{max} or J_{ss}) or the permeability coefficient (k_p) were combined with another model for estimation of percentage dermal absorption. Simple physicochemical descriptor models with MW and $LogP_{ow}$ as input parameters, complex physico-chemical descriptor models with many input parameters as well as those with quantum-chemical descriptors were included in this project. The descriptors for selected models were collected from EFSA conclusions, assessment reports and public databases or calculated with several software tools for the active substances and the co-formulants, respectively. Prediction models for dermal absorption could be successfully implemented covering two artificial neural network (ANN) models, one random walk model and one complex spreadsheet model as well as 7 algorithm-based models. The reconstruction of ANNs was realised in MATLAB 2016b using Neural Network Toolbox. All other predictions with the remaining dermal absorption models could be computed with Microsoft Excel. Multiple scenarios were accounted for during the implementation of the models: different definitions of lag-time, discrimination between experimental and computed model parameters, combination of experimental and computed parameters. This was done to generate a reasonable number of model implementation alternatives and facilitate a useful subsequent statistical analysis. Implementation of mixture models required information on the composition of the formulation which was made available by EFSA from respective sections of

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

registration reports (RRs) and Draft / Renewal Assessment Reports (DARs/RARs). However, suitable information could be retrieved only for a subset of mixtures covered in the EFSA dermal absorption dataset. In addition, information frequently related to co-formulant mixtures rather than the exact chemical composition requiring some assumptions to be made when calculating mixture descriptors.

A statistical analysis was performed addressing the relationships between predicted and measured values, as well as questions that derive from regulatory considerations. As most of the examined models yielded k_p/J_{max} values that needed to be further translated to the relevant endpoint, %DA, the DAME model (Buist 2010) was additionally applied. For the implementation of this model, information on the lag-time (t_{lag}) is required. The lag-time was either taken as 0 or was calculated; this way two sub-versions were generated for each model (except for model 7). No major variation was detected between the predicted %DA for each model sub-version, thus all further representations and analyses were performed with calculated lag-time. Another methodological issue that was addressed here is the use of information from individual replicates or from the same experimental block (blocks of replicates) for %DA predictions. In the present study, the computed %DA values were preferably derived by using, when available, as "applied concentration" the concentration that was introduced on each individual replicate (Franz cell) in one experimental block. In cases where this information was not available, the concentration means from many replicates of the same experimental block (blocks of replicates) were used. Thus, it was expected that the predictions for a set of replicates would not always be identical to one another. For this reason, the relationship between predictions for pairs of replicates from within the same block of replicates was examined, as a measure of how closely related the single replicate values are to their averaged ones.

To explore the association between measured and computed values, graphical representations with a variety of axes' scales were generated including linear, log, adjusted logit, ranks. Plots depicting the relationship between predicted and measured %DA, did not reveal any correlations between the two variables for most models. Weak to moderate associations between predicted and measured absorption, as demonstrated by the respective graphical representations and confirmed by Spearman rho and Kendall tau rank correlations analyses, were detected only for model 7 (single substance model) and for models 9 & 10 (mixture models). This finding suggested that models 7, 9 and 10 were worthy of further investigation and that the other models are less likely to be useful predictively.

For models 7, 9 and 10, a systematic search was made for subsets of the data, defined by values of covariates expected to have potential to be relevant, for which there would be a stronger correlation within the subset than for the whole data. A small number of subsets were found which had slightly stronger correlations but no subset was found which had a substantially stronger correlation.

The rate and magnitude of over- and under-prediction of absorption, and their dependence on the predicted level of absorption, were explored empirically for models 7, 9 and 10 and reported graphically and in tables.

A Bayesian random effects (BREM) statistical model of the relation between measured and predicted absorption, and incorporating random effects per active substance and block of replicate measurements was developed and used to further quantify the predictive properties of model 7. The model was used to plot the probability of a given level of over- or under-prediction as a function of the predicted level of absorption. The BREM statistical model was also used to make an adjustment to the output of model 7 to reduce the level of over- or under-prediction, depending on the predicted level of absorption.

A "Questionnaire on the practical applicability and potential regulatory implementation of *in silico* tools for the prediction of pesticides dermal absorption" was prepared considering also feedback by EFSA and the European Crop Protection Association (ECPA). The purpose of this questionnaire was to receive feedback from Member State Competent Authorities (MSCAs) on regulatory needs and raise technical questions that should be considered and addressed in the statistical review of model predictions. Thirteen (13) experts from ten (10) Member State Competent Authorities involved in the mammalian toxicology evaluation and risk assessment of active substances and their products,

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

responded to the questionnaire. In general, it was concluded that the use of *in silico* tools is currently driven by the regulatory needs as described in current legislation and related guidelines. Since *in silico* tools are currently not recommended for the prediction of dermal absorption of pesticides (EFSA, 2017), none of the experts is using them routinely. Furthermore, there is a general scepticism on the use of *in silico* tools for dermal absorption of pesticides for several reasons. Firstly, very few of the experts have the technical knowledge on how to use the available tools reliably, highlighting the need for training on the topic. In addition, the multifactorial nature of dermal absorption is recognised as a parameter limiting the *in silico* tools reliability. However, several experts would be willing to use *in silico* tools for the prediction of pesticide dermal absorption if validated methods are available and are recommended by regulatory bodies (e.g. EFSA).

Considering the result from model implementation and statistical analysis, it is recommended to further discuss the possibility of using the adjusted model 7 (NIOSH model) as a Tier 2 tool to estimate the dermal absorption value in the absence of adequate experimental data when the predicted values are in the range of 10 to 70%, with a defined level of uncertainty.

It is also recommended to further investigate the influence of physicochemical properties defined as optional by Dancik *et al.* (2013) such as pKa, water solubility, density, vehicle solubility, non-ionised unbound fractions, on the predictivity of model 7. These additional actions could be considered for potential improvement of model 7 outcome and increase predictivity.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table of contents

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

1 Introduction

1.1 Terms of Reference as provided by the requestor

This contract was awarded by EFSA to a consortium with BfR in the lead.

Members of the consortium are:

Federal Institute for Risk Assessment (BfR), Berlin, Germany

TNO innovation for life (TNO), Zeist, The Netherlands

BENAKI Phytopathological Institute (BPI), Kifissia, Athens, Greece

Durham University (DU), Department of Mathematical Sciences, Durham, United Kingdom

The Danish EPA (DK-EPA), Copenhagen, Denmark, was engaged as a sub-contractor of the consortium.

Contract title: Applicability of *in silico* tools for the prediction of dermal absorption for pesticides

Contract number: OC/EFSA/PRAS/2016/02

1.2 Interpretation of the Terms of Reference

Background as provided by EFSA

Dermal absorption is a critical element in the risk assessment of plant protection products (PPPs) for operators, workers, bystanders and residents being dermal the primary route for occupational exposure to PPPs.

Internationally agreed test guidelines exist for the performance of dermal absorption studies both *in vivo* and *in vitro*, for testing formulated product or in-use dilution. However, there is growing interest in the potential to use *in silico* (computational) tools for prediction of dermal absorption and reduce the need for testing, in particular for ethical reason.

In the last decades, non-testing approaches have been developed to predict skin absorption of chemicals. These include quantitative structure-activity relationships (QSARs), theoretical models that connect structural or physicochemical properties (descriptors) of substances to their ability to diffuse through the skin layers. In addition, more mechanistically based mathematical models of skin penetration and transport processes have been developed. A few attempts to develop models for multicomponent mixtures have been done.

Main challenges for the use of developed models based on an initial (often not homogeneous) training set of substances are their applicability to the PPPs under investigation in occupationally relevant conditions, taking into consideration also the mixture effect of formulations, and their regulatory acceptance.

From EFSA guidance on dermal absorption, the extrapolation of dermal absorption is possible only under specific circumstances (closely related formulations of the same active substance). Based on the analysis of the available approaches for predicting dermal absorption of active substances for pesticide products, QSAR models were not recommended and the need to conduct further research in the area was evidenced. Furthermore, the need for a revision of last enhancements of the *in silico* prediction tools and the evaluation of their performances using homogeneous data on PPPs has been identified by the EFSA Working Group on Dermal Absorption. The robust EFSA dataset of human *in vitro* dermal absorption studies with PPPs has been indicated as appropriate for assessing predictivity of existing *in silico* tools.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

1.3 Additional information

The overall objectives of the contract were defined as described below.

"Objective 1: provide a comprehensive and critical review of existing *in silico* models developed to predict skin absorption. The review should cover existing available models for the period from 1990 till present. The search strategy should be documented, following the criteria described in the EFSA Systematic Review Guidance, if possible. References shall be collected into EndNoteTM Library or in a format that is compatible with EndNoteTM. In the review the following information on the *in silico* models should be included: descriptors, algorithm, and training sets used for the development of the model, information on applicability domain, assessment of goodness-of-fit, robustness and predictivity, mechanistic interpretation, additional information. In addition a critical appraisal on advantages and limitations of existing *in silico* models should be provided, as well as their applicability in the regulatory context. The review should include also models for multicomponent mixtures to include vehicle effect in the prediction.

• Objective 2: compile a list of identified *in silico* models (including models for mixtures) for skin absorption prediction to be critically evaluated. The EFSA dermal absorption dataset of human *in vitro* studies with PPPs should be implemented with the descriptors to be used for testing the selected models. This information should be retrieved from available published scientific literature, chemicals databases, or calculated using software tools. In addition, to evaluate the vehicle effect, data on formulations composition of PPPs in the EFSA dermal absorption dataset will be extracted from the relevant sources (e.g. Draft Assessment Reports (DARs), Draft Registration Reports (DRRs). The evaluation should include a minimum of 10 models (at least 3 should be models for mixtures) covering skin penetration as endpoint. Moreover, a different combination of models should be also investigated as a possibility for improvement of the prediction performance.

• Objective 3: test the predictability and reliability of predictions for the identified *in silico* models against the dermal absorption dataset and provide the analysis of the outcomes including a proposal for refinement of existing models. Special attention should be given to the evaluation of the reliability and of the uncertainties of the prediction. A list of recommended models for dermal absorption prediction for pesticides should be provided. In addition, the possibility of incorporating such models into the tiered approach for the dermal exposure assessment of PPPs has to be evaluated."

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

2 Data and Methodologies

2.1 Literature Search

A search was performed in Scopus, PubMed, ToxLine and Google in order to identify publications describing mathematical/statistical models for prediction of dermal absorption across mammalian skin since 1990. The development of the search strategy is described in detail in Appendix A. In short, relevant search terms and subject areas were selected and reviewed by experts in the field. For refinement, it was investigated whether a draft search strategy developed for Scopus would reveal all 27 papers that are listed in Scopus and were reviewed in a recent publication by Buist (2016). Following revision of the initial search strategy, 26 out of 27 publications were retrieved. This was considered acceptable and the search strategy was then translated into PubMed and ToxLine search protocols. It was able to find all publications relating to dermal absorption modelling that were quoted in the recent EFSA Guidance on Dermal Absorption (2017) with the exception of two papers. One of these was not listed in the literature databases, the other was an overview paper by Mitragotri *et al.* (2011). To increase coverage, additional searches were performed including the search term "permeability", and now also the Mitragroti paper was hit. In addition, an internet search using Google was performed resulting in seven webpages containing 3 relevant hits. All search results were exported into the TNO-Literature Review Tool (TNO-LRT) and unduplicated. The Endnote™ file containing all references is available upon request at EFSA.

2.2 Literature Relevance Assessment

An initial evaluation of relevance was based on publication title. In the context of this check, "relevant papers" were defined as "papers addressing mathematical models of dermal absorption". Records considered irrelevant based on their abstract were excluded from the next selection phase. The reasons for exclusion were documented and included the following: "No mathematical model of dermal absorption", "Refers to models, but does not describe them", "PBPK model for a single substance", "Mixed model with *in silico* and experimental input" and "Other". Details of the relevance assessment and documentation are provided in Appendix B. Results of the relevance assessment for all individual references are available upon request at EFSA.

2.3 Systematic Literature Appraisal

A systematic evaluation of the scientific quality and relevance for the regulatory framework for pesticide risk assessment of the models described in the papers retrieved from public literature was performed. Proprietary *in silico* models were not excluded from the start, but evaluated along-side the non-proprietary models. Only well-defined and explicit algorithm models and reconstructable machine learning models were evaluated further. It was evaluated whether the publications provided details on the training sets used for model development and the model's applicability domain. Statistical parameters provided in the papers to illustrate the goodness-of-fit, robustness and predictivity of the models were included. This included external validations described in the original paper, not those in secondary literature. Finally, it was assessed to which extent the models offer a mechanistic interpretation of the dermal absorption process.

The first stage of the scientific review was based on answering the following questions:

1. Relevant based on complete paper?
2. Prediction for single chemical or mixture?
3. Original paper, review paper or other, non-relevant, type of paper?

Only relevant original papers were considered in the second stage of the assessment. At the second stage of the scientific review the following questions were addressed:

1. Algorithm based model?
2. Addresses a clearly defined need in pesticide RA?

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3. Are there more recent similar models by the same research group?
4. Details of the training set provided?
5. Can the model be reproduced based the materials and methods of the paper?
6. Internal validation performed?
7. External validation performed?
8. Can the model be interpreted mechanistically?
9. Proprietary model?

As soon as for a specific model either question 2 or 5 of the second stage of the scientific review was answered with "No" or question 3 with "Yes", it was not further considered in the assessment.

Quality control of the results of the systematic literature appraisal was performed by the internal reviewer. In total 26 of 129 publications describing single substance models and 5 papers describing mixture models (approx. 20% of all scientifically assessed papers) were evaluated regarding questions 1-2 and 4-9 of the second stage of scientific review. The concordance between assessor and reviewer was evaluated in a blinded manner. It was foreseen that if the percentage of publications with deviating evaluation was higher than 10%, a set of additional 20% of the papers would be evaluated by the reviewer. Any other errors would be corrected. A deviation rate of 3.2% between the internal reviewer and the initial assessor was recognized, which was regarded as acceptable according to the established protocol. Minor deviations as detected by the reviewer were adjusted after discussion with the assessor.

Further information on the methodology is provided in Appendix C.

2.4 Selection of Models for Implementation

The literature review resulted in the identification of 111 relevant single substance and mixture models, which are reconstructable, suitable for pesticide risk assessment and represent the most recent version.

A total minimum of 11 models including 3 mixture models were selected for implementation and statistical evaluation. Models predicting either maximum flux (J_{max}) or the permeability coefficient (k_p) require combination with another model for estimation of percentage dermal absorption. One of the selected models predicted percentage dermal absorption and further output parameters with corresponding entries in the EFSA dataset (also refer to section 2.7). The model selection approach can be described as follows.

In a first step, a scoring scheme based on the input parameter availability was developed to allow a ranking of single substance models (SSM) and mixture models (MM). The highest score was taken forward when more than one availability category was applicable.

In a further stage, a selection was made with the objective to ensure best possible representativeness regarding input and output parameters as well as modelling approaches, taking into account available information on existing validation efforts and training / test data set sizes.

2.5 Combination of Models

2.5.1 Sequential combination (DAME)

Since the pesticide absorption data published by EFSA only report percentage absorption and not flux or k_p as absorption parameters (EFSA, 2017), the output of all but one model needs to be converted to percentage absorption (%DA), as they predict flux or k_p. *In vitro* dermal absorption measurements of pesticides are usually executed under finite dose conditions, implying a mathematical model addressing finite dose conditions is needed for this conversion. Two publicly available finite dose models discussed by Buist, one of the authors of this report, in his 2016 thesis, were considered for this purpose: the Dermal Absorption Model for Extrapolation (DAME) (Buist *et al.*, 2010; Buist, 2016)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

and the Finite Dose Skin Permeation (FDSP) model made available on the internet by the US Centers for Disease Control and Prevention (CDC)[1]. The searches performed, which were not specifically targeted at such models, had not provided other alternatives.

DAME basically models the Franz diffusion cell and predicts *in vitro* dermal absorption from aqueous solutions, defined as the sum of the amounts of a chemical encountered in the *epidermis* and the receptor fluid. Chemical-specific inputs needed by DAME are the measured permeation constant (k_p), lag time and *stratum corneum*/water partition coefficient ($K_{SC,W}$). The permeation constant and lag time should be measured in an infinite dose *in vitro* absorption experiment. Usually, a measured $K_{SC,W}$ is not available, in which case the model uses a QSAR to calculate it. For this QSAR two additional chemical-specific parameters are needed: its MW and $logP_{ow}$. DAME can be used to convert a predicted k_p into percentage absorption by using this k_p as input instead of a measured k_p, and setting lag time to zero. The introduction of this worst case assumption (that absorption starts immediately at t = 0) is offset by also ignoring a potential skin reservoir. Alternatively, a predicted lag time can be used using the fixed relationship between lag time (t_{lag}), $K_{SC,W}$, k_p for the *stratum corneum* (k_{pSC}) and thickness of the *stratum corneum* (L_{SC}), derived by Shah et al. (1994): $t_{lag} = K_{SC,W} \times L_{SC}/6k_{pSC}$.

The FDSP model was developed by Fedorowicz *et al.* (2011) and published in peer-reviewed literature (Dancik *et al.*, 2013; Kasting and Miller, 2006; Kasting *et al.*, 2008; Miller and Kasting, 2010; Wang *et al.*, 2007). Besides predicting skin permeation under finite conditions, the FDSP model also facilitates predictions based on user specified k_p's. FDSP needs quite a number of physicochemical parameters of the penetrant as input: MW, $LogP_{ow}$, water solubility, melting point, boiling point, vapour pressure, pKa, number of rings and number of double bonds.

Table 2.1: Parameters used in DAME.

Parameter	Description	Value	Unit
MW	molecular weight	a.i.	g/mol
$LogP_{ow}$	logarithm of octanol/water partitioning coefficient	a.i.	--
Alkanol?		yes/no (a.i.)	--
k_p	predicted permeation coefficient or predicted J_{max} (mg/cm^2/h)/solubility of a.i. in the vehicle used (mg/cm^3)	0.003	cm/h
$V(SC)$	volume of Parameters used in DAME, set to a default value	0.002	mL/cm^2
$K_{SC,W}$	*stratum corneum*/water partitioning coefficient, predicted: = 0.514 $LogP_{ow}$+ 0.104 (alkanols) = 0.078 ($LogP_{ow}^2$ + 0.868 LogMW - 2.04 (non-alkanols))	a.i.	--
$V(don)$	volume of the donor cell in the predicted experiment	sp.e	mL/cm^2
t_{lag}	lag time, predicted using $t_{lag} = K_{SC,W} \times L_{SC}/6k_{pSC}$	0 or a.i.	h
L_{SC}	thickness of the *stratum corneum*, set to a default value	0.002	cm
k_{pSC}	permeation coefficient of the *stratum corneum*, approximated by using the predicted overall k(p)		cm/h
f	fraction of not-penetrated dose in donor cell: V(don) / (V(SC) x $K_{SC,W}$ + V(don))	a.i.	
$M(don, 0)$	applied dose at start of experiment	sp.e	mg/cm^2

**a.i. = specific of active ingredient, sp.e = specific for the experiment*

The performance of DAME and the FDSP model was compared by Buist (2016) using a dataset of 15 chemicals with MW varying between 60 and 391 g/mol and $LogP_{ow}$ between 0.17 and 7.7.

[1] http://www.cdc.gov/niosh/topics/skin/finiteSkinPermCalc.html

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In general, the predictions by the FDSP model did not correlate well with the measured absorption values (R^2 = 0.12) and were outperformed by the DAME predictions (R^2 = 0.64). Therefore, and because it needs less parameters as input, DAME was preferred above the FDSP model for the conversion of k_p or J_{max} in absorption percentages.

The parameters used in the DAME model to convert predicted k_p or J_{max} into percentage absorbed are listed in Table 2.1 above.

The percentage of the applied dose that had penetrated into the receptor fluid (%M(rec)) was calculated using the following formula:

$$\%M(rec) = \left\{1 - e^{\left[\frac{-f*k_p}{V_{(don)}}*(t_{(expo)}-t_{lag})\right]}\right\} * 100$$

Equation 2.5.1

in which $t_{(expo)}$ = exposure time in hours.

The percentage of the dose that had remained in the *stratum corneum* at the end of the experiment $\%M(SC)$ was calculated using the following formula:

$$\%M(SC) = [100 - \%M(rec)] * (1 - f)$$

Equation 2.5.2

The predicted (potential) absorption was subsequently calculated by adding $\%M(rec)$ and $\%M(SC)$.

2.5.2 Combinations to improve predictivity

One of the objectives of the present study was to explore the possibility to combine independent models in order to improve predictions. Accordingly, the three models showing the best individual correlation were combined in groups of two or three and predicted percentage of absorption was calculated either as the average of the two/three individual model predictions, the minimum or the maximum[2]. Predictions were made for the subset of data for which all models were applicable to avoid selection bias. Spearman rank correlations were calculated as described below.

2.6 Input Parameters related to the active substances

2.6.1 The EFSA dataset

An Excel file identified as "Human *in vitro* PPPs Dataset_CONFIDENTIAL.xlsx" was received from EFSA, which was essentially identical with the "Human in vitro dermal absorption PPPs dataset" published as supporting information to the EFSA (2017) Guidance on dermal absorption[3].

The dataset file contains information on *in vitro* dermal absorption studies performed according to OECD test guideline 428 on plant protection products and dilutions thereof. The file is composed of 6842 records/lines, with each line representing one single replicate of an experimental block. For each replicate, information on the identity and basic properties of the active substance, the origin of the data, experimental conditions and measured data is provided in the columns. The dataset provides experimental information on the dermal absorption of 193 different active substances comprising 29 different types of formulations, such as SE (Suspoemulsion, 8% of the records), WG (Water-

[2] Combination of single substance model with mixture models is allowed, since the given mixture models reflect the absorption of the a.s. and not of each single mixture component (see also 2.7.2.4).

[3] http://onlinelibrary.wiley.com/store/10.2903/j.efsa.2017.4873/asset/supinfo/efs24873-sup-0003-SupInfo_3.xlsx?v=1&s=a89137554518e850f69d836c219c853f89bd5e6c (date accessed 19/01/2018), also available at http://s3.amazonaws.com/objects.readcube.com/publishers/wiley/original_supplements/2f777ded3499aa2ccd6e3e1cabd5440ab160b9efff1ae65bf46099fc75a5dee0/efs24873-sup-0003-SupInfo_3.xlsx (date accessed 07/09/2018)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

dispersible granules, 13% of the records), EC (Emulsifiable concentrate 23%) and SC (Suspension concentrate, 30% of the records).

For the present analysis, the experimental %DA was taken as the sum of the amount directly absorbed into the receptor fluid and present in the whole skin by subtracting tape strips 1 and 2. In analogy to the statistical evaluation described in EFSA (2017) and in the interest in homogeneity of data, completion of absorption indicated by t0.5 above 75% was not considered. An extensive description of this dataset is provided in Annex A of the revised EFSA Guidance on Dermal Absorption (2017).

2.6.2 Parameters related to the active substances

For implementation of the models selected for evaluation as described in section 2.7, additional structural, quantum-chemical, and physical-chemical parameters had to be collected or calculated for the active substances listed in the EFSA dataset. A summary of the parameter values sources is provided in Table 2.2. It was noted that the same descriptor may be calculated with various tools, while the method for parameter calculation was not always clearly described in the publication and application of different methods would not necessarily result in the same parameter value.

Table 2.2: Overview of collecting source with descriptors. "Regulatory documents" refers to published EFSA conclusions, draft Registration Reports ((d)RR), DAR, Renewal Assessment Reports (RAR) and Risk Assessment Committee (RAC) opinions. "N.A." refers to not applicable.

Parameter	Experimental values source	Computed values source
MW: molecular weight	EFSA dataset	N.A.
P_{ow}: octanol-water partition coefficient	EFSA dataset	N.A.
m$LogP_{ow}$: mixture octanol-water partition coefficient (co-formulants)	Pubchem	EPI Suite
HA/mHA: hydrogen bond acceptor counts (also applies to co-formulants)	Pubchem	Molinspiration
HD: hydrogen bond donor counts	Pubchem	Molinspiration
double bonds counts	N.A.	Instant JChem (IJC)
triple bonds counts	N.A.	Instant JChem (IJC)
aromatic and non-aromatic rings counts	N.A.	Instant JChem (IJC)
TPSA/mTPSA: Topological Polar Surface Area (also applies to co-formulants)	Pubchem	Molinspiration
VP: vapour pressure	Regulatory documents	EPI Suite
MP: melting point	Regulatory documents	EPI Suite
BP: boiling point	Regulatory documents	EPI Suite
S_{aq}: water solubility	Regulatory documents	Instant JChem (IJC)
S_{oct}: olubility in octanol	Regulatory documents	N.A.
pKa	Regulatory documents	Instant JChem (IJC)
MV: molecular volume	N.A.	Molinspiration
Ve: Van der Waals volume	N.A.	Instant JChem (IJC)
EHOMO: the energy of the highest occupied molecular orbital (eV)	N.A.	MOPAC
ELUMO: the energy of the lowest unoccupied molecular orbital (eV)	N.A.	MOPAC
Qh: the sum of the net atomic charges of the hydrogen atoms bound to nitrogen or oxygen atoms	N.A.	Instant JChem (IJC)
QO.N: the sum of the absolute values of the net atomic charges of oxygen and	N.A.	Instant JChem (IJC)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Parameter	Experimental values source	Computed values source
nitrogen atoms which were hydrogen-bond acceptors		
MR: molar refractivity	N.A.	Instant JChem (IJC)
H_d or B: Hydrogen bond acceptor activity (summation)	N.A.	Abraham Descriptor Prediction
H_d or A: Hydrogen bond donor activity (summation)	N.A.	Abraham Descriptor Prediction

Experimental data on water solubility, solubility in octanol, melting point, boiling point, vapour pressure and pka values were collected, when available, from published EFSA conclusions, draft Registration Reports ((d)RR), DAR, Renewal Assessment Reports (RAR) and Risk Assessment Committee (RAC) opinions. Even though experimental conditions in *in vitro* dermal absorption studies are 32°C, data availability and comparability required to collect most physico-chemical values for (20)25 °C and pH7.

The SMILES codes of the active substances were retrieved from the internal BfR database. The SMILES codes were identified by the given active substance name and identity was furthermore verified by the given molecular weight. A cross-check was additionally conducted against displayed structures from SMILES codes listed in Pubchem public database for a few substances. If no entry was found in the internal BfR database for active substance name, a search by name was performed in Pubchem public database. In cases where no information was available in Pubchem either, the structure of those substances was clarified by google search. Initially, 198 different substance names relating to 193 active substances were provided by EFSA in the dataset within the scope of the project. During data curation, substance names were harmonized and two substances, for which the structure could not be identified unambiguously, were excluded from further analysis (refer to Appendix E for further detail). Thus, the final number of structures included in the study was reduced to 191..

The water solubility (Saq) was calculated using the proprietary software ChemAxon Instant JChem (IJC; version 17.22.0)[4] with the Solubility Calculator Plugin (version 2017). IJC calculates the intrinsic aqueous solubility of the respective substance. The calculation method is based on the article of Hou, Xia *et al.* (2004). This method uses a fragment-based method to estimate the parameter, meaning the molecule is divided into fragments and the contribution of these structural fragments to water solubility results in the estimated parameter water solubility. The function logS (logS("7.0", "mg/ml")) at pH 7.0 was used to calculate water solubility. In cases where two molecules or active substances as salt represent the active substance, water solubility was calculated for the active molecule. For example, for Aminopyralid-TIPA with SMILES code "CC(CN(CC(C)O)CC(C)O)O.C1=C(C(=C(N=C1Cl)C(=O)O)Cl)N", the water solubility was calculated only for Aminopyralid (SMILES: C1=C(C(=C(N=C1Cl)C(=O)O)Cl)N) and not for triisopropanolammonium or Aminopyralid-TIPA, representing the active substance of interest. This approach was also adopted for other active substances listed in the dataset, which consist of two molecules. Unrealistic water solubility values were obtained for some active substances, e.g. Aminoethoxyvinylglycine (Aviglycine), Trinexapac, Propamocarb HCl, Ethephon and Glyphosate reaching values of 49920.40, 5168.73, 18137.5, 143720.0 and 1931460.00 mg/L, respectively. Overall, calculated water solubility values were higher than the experimental ones. Figure 2.1 illustrates the relationship between predicted and reported Saq, which can be described mathematically by a Kendall tau value of 0.534 and Spearman rho correlation coefficient of 0.702. Even though a correlation was established, quantitative concordance between the experimental (reported) and predicted values was regarded as insufficient for the purpose of this study. It was concluded that computed water solubility should be considered of

[4] 2017, ChemAxon http://www.chemaxon.com

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

limited reliability and experimentally values for water solubility should be preferably used for model implementation.

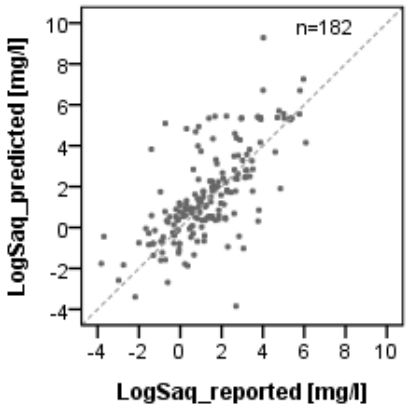


Figure 2.1: Comparison of the logarithms of the reported water solubility values to those predicted using IJC software.

The calculation of pKa values was performed with the Protonation Plugin (version 2017) in IJC and the chemical term pka ("1") for the first strongest pKa value (acidic or basic).

Where no information on melting point (MP), boiling point (BP) or vapour pressure (VP) were available in accessible EFSA conclusions, (d)RR, DAR, RAR, RAC opinions, these parameters were predicted with EPI Suite[5] using the module MPBPWIN.

The topological polar surface area (TpSA) is a descriptor associated with the polarity of the molecule, specified as the contribution of polar atoms at the surface of a molecule. TpSA values were collected from the Pubchem database, when available, and were additionally calculated with Molinspiration[6] based on SMILES codes. The calculation is in both cases based on the method described by Ertl, Rohde *et al.* (2000), where the total polar surface area is the sum of polar surface contributions of 3D generated structures of molecules.

The molecular volume (MV) was calculated with Molinspiration. SMILES codes were entered in the command line and the three dimensional volume was calculated based on a group contribution method. Geometries were optimised by semi empirical AM1 method.

The Van der Waals volume (Ve) values were calculated based on the active substances' SMILES codes and 3D geometrical conformers by using the Geometrical Descriptor Plugin of the Structural Calculations Plugin (version 2017) in IJC.

The molar refractivity (MR) was estimated based on SMILES codes by using the Refractivity Plugin of the Structural Calculations Plugin (version 2017) in IJC. The computation relies on the atomic method published by Viswanadhan, Ghose *et al.* (1989).

The number of rings was counted with the Topological Analysis Plugin of the Structural Calculations Plugin (version 2017) of IJC. Here ring count means the number of rings per molecule, which is based on the Smallest Set of Smallest Rings (SSSR) and is taken from the structure. The number of rings as required for the selected model comprises all aromatic and non-aromatic ring systems. This was covered by using the ring count function of IJC.

The number of double bonds per active substance was determined manually from structures generated from SMILES codes in IJC. All kinds of double bonds (in conjugated systems and others)

[5] US EPA. [2018]. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.

[6] free online Molinspiration property calculation service, available at: http://www.molinspiration.com/cgi-bin/properties

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

were counted as required for the respective model. A potential limitation of this procedure arises from the fact that SMILES codes represent only one mesomeric resonance structure. This may be of relevance for example for nitro groups occurring in several active substances. For such sub-structures, one double bond was counted (SMILES code: [N+](=O)[O-]).

Hydrogen bond donor HD and acceptor HA counts were registered as reported in the Pubchem public database according to the classification published by Wang, Fu *et al.* (1997). In addition, HA and HD values were also calculated with Molinspiration, where the underlying method is based on counting oxygen and nitrogen atoms for HA (nON) and OH and NH groups for HD (nOHNH).

For the implementation of one single substance model the quantum chemical descriptors Qh and QO.N were required. Qh is defined as the sum of the net atomic charges of the hydrogen atoms bound to nitrogen or oxygen atoms. QO.N relates to the sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which are hydrogen-bond acceptors. Calculation was performed using IJC including the usage of the Structural Calculations Plugin (version 2017). In a first step, SMILES codes were entered in a new substance field and hydrogen bond acceptors were afterwards indicated. Following this, the displaying of partial charges for the input molecule was selected. The descriptors were then manually calculated by summation of the respective partial charges indicated for the single atoms in the molecule. The reproduction of calculated values by Fu *et al.* (2002) was assessed for three selected substances as shown in Table 2.3. It was observed that there can be substantial deviations between Qh and QO.N values calculated in this project and those reported by Fu *et al.* (2002).

Table 2.3: Comparison of calculated and reported Qh and QO.N values for three selected substances.

Substance	Qh		QO.N	
	calculated	reported	calculated	reported
17-Oxo Dexamethasone	0.21	0.6426	1.07	1.5276
Methyl 4-hydroxybenzoate	0.29	0.2238	0.62	0.8831
3-Nitrophenol	0.29	0.2277	0.83	0.9531

The hydrogen bond donor/ acceptor activities (H_d or A and H_a or B, respectively) were estimated with a free online tool named Abraham Descriptor Prediction[7]. This model is based on the publication of Bradley et.al (2015) "Predicting Abraham model solvent coefficients". The descriptor values were predicted using a model derived from the Open Notebook Science Abraham Descriptor data. The model returned all the Abraham descriptors including the desired hydrogen bond donor activity and hydrogen bond acceptor activity. For each active substance, SMILES codes were needed for computation as well as molar volume MV $[Å^3]$ and molar refractivity $\left[10^6 * \frac{m^3}{mol}\right]$ are required input parameters and were calculated as described above. For further information on the computation of the hydrogen bond donor/ acceptor activities please refer to Appendix F.

For implementation of single substance models published by Fu, Ma and Liang (2002), EHOMO and ELUMO values were needed as input variables. For the training set used by the authors, the AM1 method was employed for calculation of these values. Accordingly, the AM1 method was also chosen to predict EHOMO and ELUMO values for the pesticidal active substances evaluated in this project. The software solution MOPAC2016[8] was preferred for this purpose as this tool is available free of charge for non-commercial purposes. SMILES were translated into MOPAC Cartesian format (MOP files) using OpenBabel software version 2.4.1 (O'Boyle *et al.*, 2011). During conversion, hydrogens were added and the 3D structure was created with coordinated centers. The AM1 Hamiltonian method was applied following geometry optimization with symmetry imposed using MOPAC2016 (32 bit-version 17.279W). In a first step, it was checked whether MOPAC2016 can reproduce the EHOMO and ELUMO values calculated by Fu, Ma and Liang (2002) to a satisfactory degree. SMILES codes were

[7] http://showme.physics.drexel.edu/onsc/models/AbrahamDescriptorsModel001.php

[8] MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

obtained for 12 substances selected from the training set used by Fu *et al.* (2002) from PubChem public database. HOMO/LUMO energies as reported by Fu *et al.* (2002) and those calculated by the method described above were not fully identical (see Appendix F). Nevertheless, for the purpose of this project, the level of reproduction was considered sufficient. The same procedure as described above was followed for predicting EHOMO and ELUMO for active substances listed in the dataset.

The collected data is provided within the respective Supporting Information (file name: "Human *in vitro* PPP EFSA dataset with add parameters and predictions.xlsx").

2.6.3 Parameters related to test item formulation

The first step to collect information on the composition of the tested formulations in the characterised study reports in the provided EFSA dataset was to collect the identifiers for the tested formulations (formulation codes) from the study reports, as far as available. Altogether, 414 different study identifiers could be found in the dataset relating to 294 study reports submitted by ECPA and 120 dermal absorption study reports evaluated by BfR. Two internal databases were searched to support the identification of the actually tested formulations in those studies. One database links the study identifier to the registration number (in Germany), while the second one links registration number to formulation codes. This allowed the collection of 142 formulation codes related to studies (study identifiers) reported by ECPA and BfR.

In some dermal absorption studies, two or more pesticide products were tested and summarised in one study report. This resulted in more formulation codes than studies. Study identifiers could be, in part, directly linked to national registration numbers for which the study was submitted. To identify the actual formulation tested from the registration number (or the product name), all listed registration numbers for a given study report were cross-checked in the database whether they match to the description of the tested formulation as in the study report.

In general, formulation codes were specific, e.g. AKD 2023, while in some reports only a universal formulation code or trade name was reported, such as Azafenidin 80 WG. Clear linking of the study report and tested formulation to the registration number or the product name was rather challenging for universal formulation codes and therefore exact matching was not possible in most cases.

The concentration of active substance in the tested product, as stated in the original study reports, was additionally entered to the extended dataset. This information helped identifying the product name. Only if the tested formulation in the dermal absorption study could be clearly matched to one registration number and product name, the information was added to the EFSA dermal absorption dataset file. Overall, it should be noted that a certain degree of uncertainty exists in identifying the tested formulation identity.

As a next step, the formulation composition was collected based on the available formulation codes. This included identity and concentration of up to ten co-formulants for each product. Considering the available information on a case-by-case basis, the main steps of the search strategy followed are summarized below:

- Search using the formulation code names available (in CIRCABC, Post Annex I data) for any relevant Registration Report (RR) available.
- If no RR was identified as relevant based on the code name provided in the EFSA dermal absorption dataset, while there is a Registration number available, a google search was conducted to obtain any other useful information (e.g. any synonymous trade names of the formulation).
- If still no RR was identified as relevant, the EFSA dms database (Active Substance Assessments Workspace) was visited to check the availability of a relevant Draft Assessment Report (DAR) or Renewal Assessment Report (RAR) and whether the formulation of interest has been evaluated in the remit of a.s. assessment.
- Finally, the BPI archive was checked.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In cases where a report (RR or DAR/RAR) was identified as relevant, a further search was conducted for the "Study identifier" in the detailed evaluation of the dermal absorption for the specific formulation. This aimed to further confirm that the composition retrieved as relevant was indeed the one of the formulation used in the dermal absorption study. It is noted, however, that in most cases the coding used for the study references was different that the one in the EFSA dataset.

When no RR was identified for the specific code name, the availability of evaluations for similar formulations was further checked. In some cases, it was found that the dermal absorption study of interest has been used for the assessment of a different formulation and thus the detailed composition of the formulation used was included in the RR of another product.

Following the search strategy described above, it has frequently not been possible to identify the formulation used in the study recorded in the database. This affected 4881 records out of 6842 in total. Records without formulation data concerned about 288 studies (140 active substances). Excluding the above studies, it was concluded that a detailed search was possible only for 140 formulations (corresponding to 1961 of 6842 records in the EFSA dermal absorption dataset provided).

Overall, the exact composition for 75 formulations was retrievable:

- for 68 formulations a relevant Registration Report (RR) – Part C has been identified in the CIRCABC,
- for 4 formulations although no relevant RR was identified in the CIRCABC it was found that the required formulation composition was included in the RAR (Volume 4) for the a.s. available in EFSA dms.
- For 3 formulations, no RR was available in any of the above data sources, but the composition details were retrieved from the relevant dossier available in BPI.

For the remaining 65 formulations, no RR/DAR/RAR was available in CIRCABC/EFSA dms which included information for the specific code. It was not possible to conclude whether any other product could be considered as relevant/similar. Detailed information on the outcome of the search for 140 formulations is available upon request at EFSA.

The final step towards the full characterization of the tested formulations was the collection or computation of physical-chemical descriptors associated with the identified co-formulants. In analogy to the active substances, SMILES codes were required for all co-formulants/co-formulant substances to further calculate physical-chemical descriptors. Notably, co-formulants frequently represented mixtures of two or more chemical substances. In addition, different co-formulant names were used across the product range to identify the same or similar (mixture of) chemical substance(s). Therefore, the following approach was chosen:

1. One CAS / SMILES / set of parameters was used when more than one chemical names were used to identify the same co-formulant having one structure and one CAS.
2. One CAS / SMILES / set of parameters was used when different CAS numbers were used for co-formulants with the same name (and identical SMILES).
3. Co-formulant entries were split into different components and information was collected for each of these components when the co-formulant name described a mixture of different components.

Grouping of co-formulants that are essentially identical reduced the number to 201 structural representations. In some cases, co-formulant codes as provided in RR and RARs did not allow to unambugiously identify the exact chemical composition. In these cases, assumptions were made based on experience and similarity to related co-formulants.

For the implementation of the selected "mixture models", information on the topological polar surface area (TpSA), hydrogen bond acceptor count (HA) and octanol-water partition coefficient ($LogP_{ow}$)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

were needed. Publicly available information on these three descriptors was collected from the Pubchem database. Whenever the co-formulants were not directly found in Pubchem with the reported names, an additional ChemIDplus search was performed with CAS to find other systematic names or synonyms. Additionally, all three descriptors were calculated with EPI Suite ($LogP_{ow}$) and Molinspiration (TpSA, HA) by using the respective SMILES codes. Further details on the computation of the physical-chemical descriptors are provided in 2.6.2.

It is to be noted that during a quality control of the database, the information on the concentration of the a.s. in the product was occasionally not identical to the information on concentration of the concentrate that was tested. This may be related to pre-application modification or experimental variability rather than data entry error.

2.7 Implementation of Models

In this chapter, short descriptions of the selected models and their corresponding implementation in the remit of the present study are outlined. It needs to be noted, that the models were not applicable for all the records/replicates provided in the EFSA dataset. This is due to restricted availability of experimental data for some records as provided in the dataset, or due to violation of models' fundamental assumptions in some cases. For example, the k_p-%DA-transformation model, DAME, is by definition not applicable to non-liquids. For more thorough explanations on the applicable records and the derived model sub-versions, please refer to Appendix G.

Predicted %DA was taken as the sum of % predicted in the receptor fluid and the *stratum corneum*, except for model 7, where the model allowed reporting of the predicted %DA as the sum of % absorbed systemically, in the *stratum corneum*, viable epidermis, and dermis.

Predicted values as well as mixture factors computed based on the information on the composition are provided within the Supporting Information (file name: "Human *in vitro* PPP EFSA dataset with add parameters and predictions.xlsx").

2.7.1 Single Substance Models

2.7.1.1 Model 1 – Frasch 2002

Model description

The Frasch model is a so called random walk model, taking into account the heterogeneous nature of the *stratum corneum*. It considers two distinct layers: the proteinaceous corneocytes and the lipid layer in between them (Frasch 2002).

Model implementation

The following formulas derived from (Frasch 2002) are used to predict k_p, expressed in cm/h:

$$k_p = k_{sc} * k_{aq}/(k_{sc} + k_{aq})$$

Equation 2.7.1

in which k_{aq} = permeation constant of the epidermis, set to 0.1151 cm/h, as postulated by Frasch (2002) and k_{SC} = permeation constant of the *stratum corneum*, expressed in cm/h, which is predicted using the following formulas:

$$k_{SC} = k_{MV} * \frac{D}{\ell^*}$$

Equation 2.7.2

in which k_{MV} = membrane (skin)/vehicle partition coefficient, D = diffusivity of the *stratum corneum*, expressed in cm^2/h, and ℓ^* = the effective path length of diffusion in the *stratum corneum*, in cm.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

k_{MV} is predicted using the following QSAR:

$$\log k_{MV} = -0.024 + 0.59 * \log P_{ow}$$

Equation 2.7.3

in which P_{ow} is the octanol/water partition coefficient of the active substance.

D is calculated using the following formula:

$$\log D = \frac{\log D_{COR}/D_{LIP}}{1 + e^{-\frac{\log K_{COR/LIP} - 0.1974 + 0.3668 * \log D_{COR}/D_{LIP}}{0.2488 - 0.134 * \log D_{COR}/D_{LIP}}}} + \log 3.6 * 10^{-5}$$

Equation 2.7.4

in which D_{COR} = the diffusivity of the corneocytes, in cm^2/h; D_{LIP} = the diffusivity of the lipid layer of the *stratum corneum*, in cm^2/h.

$K_{COR/LIP}$ is the corneocyte/lipid layer partition coefficient, which is predicted using the following QSAR:

$$\log K_{COR/LIP} = 2.4194 - 0.8075 * \log P_{ow}$$

Equation 2.7.5

Also $\log D_{COR}/D_{LIP}$ is predicted with a QSAR:

$$\log D_{COR}/D_{LIP} = -0.0087 * \mathrm{MW}$$

Equation 2.7.6

in which MW is the molecular weight of the active substance.

To conclude, the effective path length (ℓ^*) is calculated using the following formula:

$$\ell^* = 0.003 * [1 - 0.9113 * \log K_{COR/LIP} + 0.9896 * (\log K_{COR/LIP})^2 + 0.3111 * (\log K_{COR/LIP})^3]$$

Equation 2.7.7

For lipophilicity ($LogP_{ow}$) and molecular weight (MW) experimental values were used, as provided by the EFSA dataset. The permeability coefficient k_p was then transformed to %DA by feeding the predicted k_p values to the DAME model (Buist *et al.* 2010 and Buist 2016). For implementation of the DAME model please refer to section 2.5.1.

2.7.1.2 Model 2 – Potts & Guy 1992

Model Description

The classical k_p prediction model of Potts and Guy (Potts & Guy, 1992) was developed based on measured k_p values collected by Flynn (1990).

Model Implementation

The model was used by implementing the formula provided in the original publication:

$$\log k_p \left[\frac{cm}{s}\right] = -6.3 + 0.71 * \log P_{ow} - 0.0061 * MW \left[\frac{g}{mol}\right]$$

Equation 2.7.8

The algorithm described in Equation 2.7.8 was implemented in a separate Excel sheet. For lipophilicity ($LogP_{ow}$) and molecular weight (MW) experimental values were used, as provided by the EFSA dataset described in section 2.6.1. The permeability coefficient k_p was converted to cm/h by multiplying it with 3600 and then transformed to %DA by feeding the predicted k_p values to the DAME model (Buist *et al.* 2010 and Buist 2016). For implementation of the DAME model please refer to section 2.5.1. This

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

additional computation step yielded two results since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

2.7.1.3 Model 3 – Magnusson 2004

Model Description

A collection of published data for human skin epidermal permeation was used to define the relationship between the solute J_{max} and solute physical-chemical properties. First, an algorithm was developed with training data on experimental maximum flux from aqueous solutions and it was validated with experimental data for full- and split-thickness skin, ionized solutes, pure solutes and solutes in propylene glycol vehicle. Linear regression revealed that the molecular weight (MW) was the dominant determinant for the training set under consideration (n=87). Stepwise inclusion of additional parameters such as solubility in octanol, melting point, and hydrogen bond acceptor capability (also known as hydrogen bond acceptor activity H_a or B) only marginally improved the regression. After validating the initially derived equation with the subsets described above (r^2 ranging from 0.282 to 0.784), linear regression was performed with the training and validation datasets merged together (n=278) to derive the final equation:

$$\log J_{max}\left[\frac{mol}{cm^2 * h}\right] = -4.52 - 0.0141 * MW\left[\frac{g}{mol}\right], r^2 = 0.688$$

Equation 2.7.9

Model Implementation

In the present study, Equation 2.7.9 was used to calculate the maximum flux J_{max}. In order to obtain %DA values, the maximum flux J_{max} needed to be transformed to k_p. This was done by dividing the $log\, J_{max}$ by the water solubility S_{aq} according to the following formula referenced in the paper:

$$k_p = \frac{J_{max}}{S_{aq}}$$

Equation 2.7.10

Where k_p is expressed in cm/h , J_{max} in $\frac{mol}{cm^2 * h}$, S_{aq} in mol/l.

Since S_{aq} is captured in the database in mg/l , it needs to be converted to S_{aq} in mol/l according to the following calculation:

$$S_{aq}\,[mol/l] = \frac{S_{aq}\,[mg/l]}{\mathrm{MW}\,[g/mol]} * 10^{-3}$$

Equation 2.7.11

Consequently, the following formula was used for the final calculation of k_p:

$$k_p = \frac{J_{max} * MW}{S_{aq}}\left[\frac{\cancel{mol} * \cancel{g} * l}{cm^2 * h * \cancel{mol} * m\cancel{g}}\right] = \frac{J_{max} * MW}{S_{aq}}\left[\frac{\cancel{cm^2} * cm}{\cancel{cm^2} * h * 10^{-3}}\right] = \frac{J_{max} * MW}{S_{aq}} * 1000\left[\frac{cm}{h}\right]$$

Equation 2.7.12

With S_{aq} corresponding to the one in the database expressed in mg/l.

Experimental values were used for molecular weight (MW) throughout all calculation steps. For water solubility ($S_{aq}\,[mg/l]$) experimental data from publicly available regulatory documents - EFSA conclusion, draft Registration Report ((d)RR), Draft Assessment Report (DAR), Renewal Assessment Report (RAR) and Risk Assessment Committee (RAC) Opinion - were used, where available. It was noticed that the retrieved experimental data on the water solubility are not covering all substances in the scope of the study. For that reason, the whole simulation was rerun with solely computed data

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

derived from Instant JChem (for methodology see section 2.6.2). This way, two different k_p values were computed for each record, one solely based on experimental input parameters (MW, S_{aq}) and one with experimental MW and computed S_{aq}. The rationale behind this decision is that this way evaluation of the model is possible under two scenarios: when experimental physical chemical values are readily available and when not.

By translating k_p into %DA, the two different k_p values were fed into a model where two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated), which in turn yielded four different %DA values. For more detailed explanation of the k_p-%DA transformation please refer to the description of the DAME model in section 2.5.1.

2.7.1.4 Model 4 – Fu 2002 Algorithmic solution

Model Description

The Abraham dataset (Abraham *et al.*, 1997) was used to construct two skin permeability models: an algorithm and an Artificial Neural Network (ANN) model (ANN explained in 2.7.1.5). Both models are based on the same input parameters:

i. *MV*, which is the molecular volume and is given in nm^3 ($1\text{Å}^3 = 10^{-3} nm^3$)
ii. Q_H, which is the sum of net atomic charges of hydrogen atoms attached to oxygen or nitrogen atoms
iii. $Q_{O,N}$, which is the sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which were hydrogen bond acceptors
iv. E_{HOMO}, which is the energy of the highest occupied molecular orbital and is expressed in eV
v. E_{LUMO}, which is the energy of the lowest unoccupied molecular orbital and is expressed in eV.

The algorithmic model was developed by simple linear regression and is defined as follows:

$$\log k_p \left[\frac{cm}{s}\right] = -4.762 + 3.691 * 10^{-3} * MV - 2.856 * Q_H - 2.194 * Q_{O,N} - 0.03267 * E_{HOMO} - 0.2196 * E_{LUMO}$$

Equation 2.7.13

Model Implementation

In the present study, the algorithmic model was implemented according to Equation 2.7.13 in a separate Excel sheet added to the EFSA dataset file. For the description of the computation of the necessary parameters for model implementation please refer to section 2.6.2.

Similarly to the other models, the calculated k_p was translated to %DA using DAME as described in section 2.5.1, yielding two different %DA values, since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

2.7.1.5 Model 5 – Fu 2002 Artificial Neural Network (ANN)

Model Description

The Fu *et al.* (2002) Artificial Neural Network (ANN) model is a four layer network that includes an input layer, two hidden layers, and an output layer. The description of the input layer (input parameters) is provided in section 2.7.1.4. The two hidden layers both consist of 4 neurons and the output layer of a single neuron, the permeability coefficient $\log k_p$. The ANN was initially trained with 45 input-output datasets for 100000 cycles and was validated by prediction of the $\log k_p$ values of 8 additional compounds outside of the training set (mean prediction error 2.6%). In comparison to the respective algorithmic regression model (see 2.7.1.4) the ANN showed better predictive capacity.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Model Reconstruction and Implementation

The feedforward artificial neural network (ANN) model was reconstructed on the basis of the training set published in the paper. The reconstruction was realized in MATLAB 2016b using Neural Network Toolbox, with 5 inputs, 20 hidden layers, and an output. The inputs and outputs correspond to the same variables used in the Fu ANN model. The correlation between observed and calculated k_p using the original and the reconstructed ANN model is high (r^2 = 0.99 in both cases) and the concordance is good (RMSE for log k_p = 0.12 and 0.10, respectively). Also the concordance between the predictions by Fu *et al.* and the reconstructed model is good (RMSE = 0.12). While the Fu *et al.* model provides a good prediction of the test set (r^2= 0.98, RMSE = 0.16), the reconstructed model performs less well (r^2= 0.79, RMSE = 0.69). This is mainly due to one outlier (corticosterone) (see Figure 2.2).

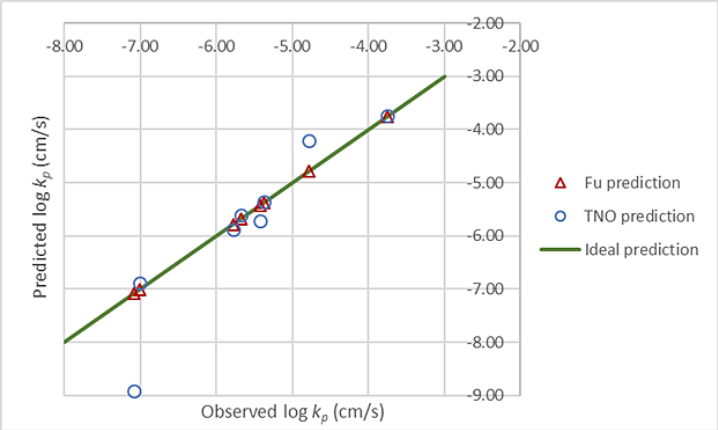


Figure 2.2: Prediction of the test data set by Fu *et al.* (2002) and the reconstructed ANN.

This trained ANN model was further used to estimate the value of $\log k_p$ given the input values: MV, Qh, QO.N, EHOMO, and ELUMO for 190 sets of a.s. data from the EFSA dataset. Aminopyralid-Olamine corresponding to 24 individual records was excluded. For molecular volume, the van der Waals volume was used.

Similarly to the other models, the calculated k_p was translated to %DA using DAME as described in section 2.5.1, yielding two different %DA values, since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

2.7.1.6 Model 6 – Milewski 2012

Model Description

The starting point for the development of the Milewski et. al (2012) model was Fick's first law of diffusion, where, for aqueous formulations, the maximum transdermal flux $\log J_{max}$ can be expressed as:

$$\log J_{max} = \log k_p + \log S_{aq}$$

Equation 2.7.14

By using the "classical" Potts & Guy (1992) algorithm for estimation of $\log k_p$[9] and the general solubility equation by Yalkowsky and Valvani (1980)[10], a generic algorithm that relates $\log J_{max}$ with lipophilicity $\log \mathrm{K_{ow}}$, molecular weight MW, and melting point MP was derived:

[9] $\log k_p = -2.72 + 0.71 * \log \mathrm{K_{ow}} - 0.0061 * MW\ [cm/h]$, see also Equation 2.7.8 (transformed)

[10] $\log S_{aq} = -0.0102 * (MP - 25) - 1.031 * \log \mathrm{K_{ow}} + 0.424\ [mol/l]$

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

$$\log J_{max}\left[\frac{nmol}{cm^2 * h}\right] = a + b * \log \mathrm{P_{ow}} + c * MW + d * (MP - 25)$$

Equation 2.7.15

Equation 2.7.15 served as a basis for multiple linear regression with the training set (n=87) described in Magnusson *et al.* (2004) (see also 2.7.1.3). The training dataset comprised compounds spanning over a broad range of physical-chemical properties: $\log \mathrm{K_{ow}}$ (-4.67 to 4.52), MW (18 to 477) and MP (for liquids to 293 °C). The derived model (r^2=0.90) was validated with the validation set (n=121) described by Magnusson *et al.* (2004) (see also 2.7.1.3). The training and validation datasets were merged and were used to derive the final algorithm by linear regression analysis:

$$\log J_{max}\left[\frac{nmol}{cm^2 * h}\right] = 4.6 - 0.219 * \log \mathrm{K_{ow}} - 0.0086 * MW\left[\frac{g}{mol}\right] - 0.0102 * (MP - 25), \mathrm{r^2} = 0.81$$

Equation 2.7.16

Model Implementation

Equation 2.7.16 was applied to derive values for the maximum flux, where, by definition of the model, MP is to be set at 25°C for liquid permeants. Differentiation between liquid and non-liquid permeants was made by consulting the definitions in Table B.2 of the EFSA (2017) Guidance on dermal absorption. For WP, WG, WDG, SG, SP, GEL, a differentiation between liquid and non-liquid was done by relating the formulation type to the tested concentration (concentrate/dilution). For these formulation types the concentrate was taken as non-liquid and the dilutions thereof as liquid. For all other formulation types that are identified in the table as non-liquids (RB, Pellet Bait, Wax Block, Pasta Bait, Pellets, PS, DS) no such differentiation was made, since these are not to be further diluted. This resulted into the identification of 6249 cases as liquids and 517 as non-liquids and 42 as "not-clear".

Setting the melting point at 25°C for liquid permeants results into not taking this parameter into consideration for the calculation of $\log J_{max}$, meaning that Equation 2.7.16 above remains as it is for non-liquids and becomes for liquids:

$$\log J_{max}\left[\frac{nmol}{cm^2 * h}\right] = 4.6 - 0.219 * \log \mathrm{K_{ow}} - 0.0086 * MW\left[\frac{g}{mol}\right]$$

Equation 2.7.17

Experimental values, as registered in the EFSA dataset, were used for molecular weight MW and lipophilicity $\log \mathrm{K_{ow}}$ throughout all calculation steps. Experimental values were used for the melting point MP, when available, as retrieved from publicly available regulatory documents - EFSA conclusion, draft Registration Report ((d)RR), Draft Assessment Report (DAR), Renewal Assessment Report (RAR) and Risk Assessment Committee (RAC) Opinion. In the absence of experimental data for the melting point – corresponding to 2.3% of the total data under evaluation, computed data from EPI Suite were considered (see Table 2.4).

Similarly to the Magnusson model, $\log J_{max}$ had to be converted into k_p by dividing the $log\, J_{max}$ by the water solubility S_{aq} (see section 2.7.1.3), and consequently k_p had to be transformed to %DA via the DAME model. This, as also described in section 2.7.1.3, yielded in total four different %DA values: two different k_p values for each record, one solely based on experimental input values for S_{aq} and one on computed S_{aq}, combined with two scenarios for t_{lag} in DAME (refer to section 2.5.1 for description of DAME model).

Considering the fact that, by definition, DAME is applicable only to liquid permeants, all non-liquid $\log J_{max}$ (and therefore k_p) values that were calculated by Equation 2.7.16 were not accounted for in the final calculations of %DA. Thus, the absorption values presented for this model combination (Milewski-DAME) are based on $\log J_{max}$ / k_p values that are derived by Equation 2.7.17, so that in the end the contributions of the melting point are not taken into account for the predictions. To explore

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

the role of the melting point for the predictions on the EFSA dataset, the model was rerun based on Equation 2.7.16 without distinguishing between liquids and non-liquids.

Table 2.4: List of a.s. for which no experimental information were available on the Melting Point and computed values form EPI Suite were used.

Active substance name	Melting Point [°C] (experimental, public source)	Melting Point [°C] (EPI Suite)
2,4-D EHE (2-ethylhexyl 2-(2,4-dichlorophenoxy)acetate)	NA	121.30
Flupyrsulfuron methyl	NA	248.61
Indaziflam	NA	177.65
Iofensulfuron-sodium	NA	330.76
Metosulam	NA	232.08
Quinoxyfen	NA	148.09
Triclopyr (TEA)	NA	135.20

2.7.1.7 Model 7 – Dancik / NIOSH model

Model Description

The NIOSH model described by Dancik *et al.* (2013) is a transient model simulating bioavailability of chemicals from dermal exposure. The skin is treated as a slab of three different components, for which various kinetic parameters are estimated: the *stratum corneum* (SC), the viable epidermis (VE) and the dermis. The SC can be either partially or fully hydrated. The SC follows the two dimensional brick and mortar structure, which is composed of lipid and aqueous phases, along with corneocytes, which are composed of proteinic and aqueous phases.

The model is a spreadsheet-based model developed by the University of Cincinnati and The National Institute for Occupational Safety and Health (NIOSH) of the U.S. Centers for Disease Control and Prevention (CDC). In addition, a Java tool was created and made available as "The Finite Dose Skin Permeation Calculator"[11].

Model Implementation

It was not possible to execute the java applet provided on the website. However, the original Excel™ spreadsheet and associated add-ins identified as "Finite Dose Skin Permeation Calculator"[12], along with all the necessary files and an instruction sheet to run the code in the Excel/VB format was provided by the University of Cincinnati. The batch version of the Calculator[13] was also provided by the model creators and used for final predictions.

Subset selection

Implementation of this model proved far more time consuming than for the other models. Therefore, it was decided to apply the model only on a subset consisting of 48 out of the 191 available pesticide active substances within the EFSA dataset. The selected subset is the result of the integration of three smaller subsets which were derived based on the following criteria:

SUBSET 1 (14 active substances): Priority was given to substances for which experimental data were retrievable for all the model input parameters (see chapter 2.6.2). This applies to the following physical-chemical descriptors: No. of rings, No. of double bonds, Molecular Weight (g/Mol), $LogP_{o/w}$,

[11] https://www.cdc.gov/niosh/topics/skin/finiteskinpermcalc.html

[12] 4-Case 3-Layer (4C3L) Diffusion Model Parameters, Version: 2.1, Spreadsheet update: 20/8/2014

[13] 4C3L M.A. Miller/G.B. Kasting, Spreadsheet update: 22/7/2015

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Melting Point [°C], Boiling Point [°C], Vapour Pressure [Pa at 25°C], pKa at 20 °C, Water solubility [mg/L at pH 7 and 20/25 °C], Solubility in octanol [g/L at 20/25 °C]. This criterion was set in order to minimise the use of calculated values and therefore the risk of introducing additional uncertainty in the computation.

SUBSET 2 (9 active substances): Another criterion for the selection of substances was the availability of data on formulation composition, as retrieved during the data collection process on co-formulants. The idea behind this approach was to allow comparison of the predicted values from the NIOSH model (single substance model accounting for some vehicle effects) to the predictions made using mixture models. Note that two of such substances (cyflufenamid and cyprodinil) were already included in SUBSET 1.

SUBSET 3 (25 active substances): From the remaining 170 a.s.[14], another 25 a.s. were selected after considering the upper and lowest extremes (minimum/maximum) and medians of all relevant input parameters.

Adequate representativeness of the final combined subset (SUBSET 1 & SUBSET 2 & SUBSET 3) in comparison to the total EFSA dataset (191 a.s.) was verified graphically (histograms) and via descriptive statistics.

Implementation

For efficiency reasons, the batch version of the Calculator was used when a group of substances was tested simultaneously. Input data were added for each compound in the following spreadsheets: "Phys", "struct", "dose", "environ", "misc".

The "Phys" spreadsheet contained information on the Physical Properties of each substance:

4-Case 3-Layer Diffusion Model Parameters						M.A. Miller/G.B. Kasting								
Physical Properties				T_s =	32	°C								
					T	$calcP_{vp}$	Compound			mT	Immobile Vehicle Properties			
Compound		log K_{oct}	$calcP_{vp}$		of	@ Temp	MP	Density	mS_w	S_w	Aqueous?	Density	Sv	
Name	Abbrev		@ T_s	mP_{vp}	mP_{vp}	of mP_{vp}	°C	g/cm³	g/L	°C	Name	g/cm³	g/L	pH
cymoxanil	1048	0.67	0.00E+00			1.13E-06			0.78	20	water	TRUE	1	
cymoxanil	1049	0.67	0.00E+00			1.13E-06			0.78	20	water	TRUE	1	
cymoxanil	1050	0.67	0.00E+00			1.13E-06			0.78	20	water	TRUE	1	
cymoxanil	1051	0.67	0.00E+00			1.13E-06			0.78	20	water	TRUE	1	
cymoxanil	1052	0.67	0.00E+00			1.13E-06			0.78	20	water	TRUE	1	
cymoxanil	1053	0.67	0.00E+00			1.13E-06			0.78	20	water	TRUE	1	
Cyprodinil	1253	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1254	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1255	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1256	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1257	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1258	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1259	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1260	4	0.00E+00			3.53E-06			0.013	20	octanol	FALSE	0.83	140
Cyprodinil	1261	4	0.00E+00			3.53E-06			0.013	20	water	TRUE	1	

Figure 2.3: Example of completion of the "phys" spreadsheet.

[14] Total - SUBSET 1 - SUBSET 2 + 2 common a.s. from SUBSET 1 & SUBSET 2 = 191-14-9+2 = 170 a.s.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The "struct" spreadsheet contained information on the Structural Properties of each substance:

4-Case 3-Layer Diffusion Model Parameters																
Structural Properties			C11 H10 O3												I don't know the molecular formula **or** there are other atoms not listed here.	Pharmacophore?
Compound		MW												Ring		
Name	Abbrev	g/mol	C	H	N	O	S	Br	Cl	F	I	P	=	≡	Systems		
Amisulbrom	271	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	272	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	273	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	274	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	275	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	276	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	277	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	278	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	279	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	280	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	281	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	282	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	283	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	284	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	285	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE
Amisulbrom	286	466.3	13	13	5	4	2	1		1			10	0	3	FALSE	FALSE

Figure 2.4: Example of completion of the "struct" spreadsheet.

The "dose" spreadsheet contained information on the Dose / Regimen of each substance:

4-Case 3-Layer Diffusion Model Parameters			
Dose/Regimen		Amount of Compound Applied	Immobile Vehicle Applied
Compound Name	Abbrev	µg/cm^2	mg/cm^2
Boscalid	470	11.3	9.99
Boscalid	471	11.3	9.99
Boscalid	472	11.3	9.99
Boscalid	473	11.3	9.99
Boscalid	474	11.3	9.99
Bromoxynil phenol	511	2507	7.52
Bromoxynil phenol	512	2507	7.52
Bromoxynil phenol	513	2507	7.52
Bromoxynil phenol	514	2507	7.52
Bromoxynil phenol	515	2507	7.52
Bromoxynil phenol	516	4.72	10.04
Bromoxynil phenol	517	4.72	10.04
Bromoxynil phenol	518	4.72	10.04
Bromoxynil phenol	519	4.72	10.04
Chlorantraniliprole	585	1862	7.45
Chlorantraniliprole	586	1862	7.45
Chlorantraniliprole	587	1862	7.45
Chlorantraniliprole	588	1862	7.45
Chlorantraniliprole	589	1862	7.45
Chlorantraniliprole	590	1862	7.45
Chlorantraniliprole	591	4.48	5.97

Figure 2.5: Example of completion of the "dose" spreadsheet.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The "environ" spreadsheet contained information on the Skin Properties:

4-Case 3-Layer Diffusion Model Parameters			M.A. Miller/G.B. Kasting	22/7/2015	version
Skin Properties					
Species:	Click here for default thicknesses		**Output Parameters**		
human			maximum duration of simulation	24	hours
Hydration State:			set concentration	enter on "misc" sheet	
fully hydrated		pH	☑ Flux Plot		
☑ stratum corneum	43.4 µm	5.0	☑ Cumulative Amount Plot		
☑ viable epidermis	100 µm	7.4	☑ Concentration Profiles		
☑ dermis	300 µm	7.4	minimum output step size	10	minutes
in vivo or *in vitro* ?			maximum step size	5	hours
in vitro					
			Environmental Parameters		
For variable SC diffusivity, see the SC sheet.			Surface Temperature	32	°C
			Wind Velocity	0.17	m/s
			occlusion		not active yet

Figure 2.6: Example of completion of the "environ" spreadsheet.

The "misc" spreadsheet contained Miscellaneous information on each substance:

4-Case 3-Layer Diffusion Model Parameters				
Compounds	CAS number	Formula	Class	SMILEs
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Amisulbrom	348635-87-0	C13 H13 N5 O4 S2 Br F	Pesticide	CN(C)S(=O)(=O)n1cnc(n1)S(=O)(=O)n1c(C)c(Br)c2ccc(F)cc12
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Azoxystrobin	131860-33-36	C22 H17 N3 O5	Pesticide	CO\C=C(/C(=O)OC)c1ccccc1Oc1cc(Oc2ccccc2C#N)ncn1
Boscalid	188425-85-20	C18 H12 N2 O Cl16	Pesticide	Clc1ccc(cc1)-c1ccccc1NC(=O)c1cccnc1Cl
Boscalid	188425-85-20	C18 H12 N2 O Cl16	Pesticide	Clc1ccc(cc1)-c1ccccc1NC(=O)c1cccnc1Cl
Boscalid	188425-85-20	C18 H12 N2 O Cl16	Pesticide	Clc1ccc(cc1)-c1ccccc1NC(=O)c1cccnc1Cl

Figure 2.7: Example of completion of the "misc" spreadsheet.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Special considerations were made on the following areas:

a) Vehicle settings

Although the NIOSH model is a simple model, information on the vehicle is still required. For all dilutions, the vehicle was assumed to be water. For concentrates, vehicle selection was dependent on the formulation category. The different formulation types considered in this project were grouped into four broad categories (Table 2.5) as recommended in table B.2 of the EFSA Guidance on dermal absorption (EFSA, 2017):

Table 2.5: Product formulation and selection of vehicle and vehicle properties for implementation of model 7.

Formulation categories	Formulation type*	Vehicle	Volatility	Mobile / Immobile
(1) organic solvent-based	EC, EW, SE, OD, ME	Octanol	Non-volatile	Immobile
(2) water-based	SL, SC, FS, FL	Water	Non-volatile	Immobile
(3) solid	WP, WG, SG, SP	Water	Non-volatile	Immobile
(4) other	CS	Water	Non-volatile	Immobile
	CB, RB, WB, PB, P	Neat	NA	NA
	Al, NS	NA	NA	NA

* *Abbreviations as indicated in the EFSA Guidance on dermal absorption (EFSA, 2017): Emulsifiable concentrate, EW: Emulsion/oil in water, SE: Suspo-emulsion, OD: Oil-based suspension concentrate, ME: Microemulsion, SL: Soluble concentrate, FS: Flowable concentrate for seed treatment, FL: Flowable, WP: Wettable powder, WG/WDG: Water-dispersible granules, SG: Water-soluble granules, SP: Water-soluble powder, DS: Powder for dry seed treatment, CS: Capsule suspension.*
Additional abbreviations not included in the EFSA Guidance on dermal absorption (EFSA 2017): Al: Experimental solution of active substance in solvent CB: Bait concentrate, NS: Not specified, RB: Bait/ready for use, WB: Wax block, PB: Pasta Bait, P: Pellets. NA: not applicable.Additional formulation types that did not fit in Categories 1-3 were assigned in Category 4.

For vehicle selection, it was considered that the vehicle could be either immobile or highly volatile according to the definitions described by Dancik *et al.* (2013):

"A vehicle containing immobile components (i.e., none of it is depleted from the skin surface by either evaporation or penetration into the skin) or highly volatile components (which evaporate instantaneously) may be simulated. When deciding on whether a vehicle component is best modelled as immobile or highly volatile, the user should not only consider the physical and structural properties of the component, but also the degree to which the donor solution is occluded, the amount of airflow over it, and the volume of the headspace. Partial occlusion affects the evaporation of the permeant as well, and the wind velocity should be appropriately adjusted in that case."

Considering these definitions, the following assumptions were made for vehicle selection:

- *Category 1 - organic solvent-based formulations*

Octanol was selected as an appropriate vehicle, since measured data for the solubility in octanol were available for the selected a.s.. Although octanol is ordinarily volatile (VP=8.7 Pa at 20°C), it was assumed to be an immobile vehicle, since evaporation is typically prevented by occlusion. In this case, as described by Dancik *et al.* (2013), the vehicle/water partition coefficient ($K_{v/w}$) is estimated by considering the ratio of solubility in octanol divided by the solubility in water. No correction for vehicle ionisation is performed in the model and thus no details on pH are required. Therefore, for organic solvent-based formulations an occluded system and fully hydrated skin conditions were considered.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

- *Category 2 - water-based formulations*

It is reasonable to assume that the vehicle is primarily water. Dancik *et al.* (2013) suggest that water can be either considered a volatile or non-volatile vehicle. Under occlusive conditions (immobile vehicle) water may be taken as non-volatile since evaporation is prevented by occlusion. In this case, the non-ionised permeant in the vehicle is estimated and the vehicle/water partition co-efficient ($K_{v/w}$) is corrected for ionisation (Dancik *et al.,* 2013).

In any case, under occlusive conditions, water volatility has low impact. As a trial, the NIOSH model was run for azoxystrobin SC formulation (concentrate) under occlusive conditions considering water to be (i) volatile and (ii) non-volatile. The results obtained were comparable with predictions for Maximum Absorptive Flux (µg/cm^2 x h) of 0.00385 and 0.00383, and for Systemically Absorbed (%) of 0.00673 and 0.00673, respectively. Thus, for water-based formulations an occluded system and fully hydrated skin conditions were considered. Water was assumed to be non-volatile.

- *Category 3 - solid formulations*

Dancik *et al.* (2013) state that *"the permeant may be applied neat or dissolved in a vehicle"*. For solid formulations, three different options may apply:

Option 1 - assume "neat" compound, without any vehicle, or

Option 2 - assume that the water used for wetting of the solid formulation in accordance with OECD TG 428 can dissolve the permeant.

Option 3 - treat as not applicable in results sheet (label NA).

Although Option 3 would have been consistent with the approach taken in the implementation of some other models for solid formulations, it was considered more relevant to select between Options 1 and 2. In order to make a justified decision, the EFSA Guidance on dermal absorption (EFSA, 2017) was consulted. In the Guidance, the following is recommended:

"Solid material should be moistened with a minimal volume of vehicle (e.g. water or physiological saline) to make a paste. This is to mimic sweat on the skin or occlusive conditions under clothing. Since dermal exposure to granular products is usually in the form of dust, the granules should be ground and moistened before application to the skin. Organic solvents should not normally be used (see EFSA PPR Panel, 2011, section 3.2.) (If solids are not moistened then the validity of the study is questionable. If solids are not moistened but occlusive conditions are used then the study can be considered a reasonable match to actual exposures, except for granules.)"

Considering the above, it is reasonable to assume that either (1) there is no vehicle (neat), but occlusive conditions apply, or (2) the vehicle is water which mimics occlusive conditions under clothing. Therefore, in case of solid formulations included in category 3, the vehicle was taken as water, since it is possible to make a paste with it. An occluded system and fully hydrated skin conditions were considered.

- *Category 4 - other formulations*

For CS formulations, the vehicle was water, an occluded system and fully hydrated skin conditions were considered. This assumption was based on the EFSA Guidance on dermal absorption (EFSA, 2017), where a CS formulation is defined as *"a stable suspension of microencapsulated active substance in an aqueous continuous phase, intended for dilution with water before use"*.

When a vehicle could not be specified based on the formulation type (e.g. RB, WB, PB, P), the a.s. was assumed to be "neat". In the absence of a relevant vehicle, the model could not run for these mixtures and NA was indicated in the results spreadsheet. When there was limited or no information on the formulation type (e.g. NS, AI), it was not possible to select an appropriate vehicle and no predictions were made for these concentrates. These specific cases of a.s. in concentrates where implementation of the NIOSH model was not possible are summarised in Table 2.6.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 2.6: Active substances for which it was not possible to run the NIOSH model for concentrate formulations. N.A. refers to not applicable.

Active substance name	Formulation type	Vehicle	Reason for not running the NIOSH model
Boscalid	NS	NA	Unknown formulation type
Cholecalciferol	RB	Neat	"Neat" compound - RB formulation is designated to attract and be eaten by the target pests; vehicle is not relevant cannot be specified.
Difenacoum	Pellet Bait Wax Block Pasta Bait	Neat	Formulation types are designated to attract and be eaten by the target pests; vehicle is not relevant cannot be specified.
Alpha-cypermethrin	CB	Neat	CB is a bait concentrate and can be either a solid or a liquid intended for dilution before use as bait; the physical state of the formulation was not known and it was not possible to specify a vehicle.
Metaldehyde	RB Pellet Bait	Neat	Formulation types are designated to attract and be eaten by the target pests; vehicle is not relevant cannot be specified.
Oxydemeton-methyl	AI	N.A.	Experimental solution of active substance in solvent; the vehicle can be organic or water-based. These solutions are not comparable with organic or water-based formulations.

b) Observations during data entry

- **Use of input parameters:** The necessary input parameters for the implementation of the model are: $LogP_{ow}$, MW, VP, MP, BP, no. of double bonds, no. of triple bonds, no. of aromatic and non-aromatic rings. In Dancik *et al.* (2013) it is stated that use of some optional parameters is also possible (ρ, Saq, measured k_p, pKa). Here, the number of input parameters was kept to the minimum necessary ones. It deemed more realistic that, in case the model is to be implementated in the future in the regulatory context, optional parameters would not be used consistently for a variety of reasons including lack of availability. Since there is no clear indication in the relevant publication that the inclusion of the optional parameters would lead to improved results, it seemed acceptable not to over-complicate the implementation with additional parameters.
- **Exposure time**:The exposure time (8 or 10 or 12 or 24h) was indicated in a separate box in the "dose" sheet:

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Compound	Immobile			
Multiple	Vehicle			Explanation of
Application	Application	Time,	Removal	Removal Scenario
Dose, µg/c	Dose, mg/	hours	Scenario	numbers
Default	Default	Default	0	0
0	0.00	10	1	1
Default	Default	Default	0	2
Default	Default	Default	0	3
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	
Default	Default	Default	0	

Figure 2.8: Example for setting time in the "dose" sheet. In this example, the test substance was fully removed (removal scenario 1) after 10 hours exposure.

It is noted that there are 3 removal scenarios in the Calculator. In removal Scenario 1 all of the permeant and the vehicle are removed from the surface. In removal Scenarios 2 and 3 some of the permeant and the vehicle are removed from the surface, but none is reapplied; the amount removed should be indicated. For the purposes of this project, only Scenario 1 was considered relevant.

- **Skin properties.** Regarding skin thickness, default values were selected in the "environ" sheet as follows:
  - species = human
  - hydration state = fully hydrated
  - *in vivo* or *in vitro* = *in vitro*
  - thickness and pH of *stratum corneum* = set to default
  - thickness and pH of viable epidermis = set to default
  - thickness and pH of dermis = based on the "Skin type" column of the dataset file as follows: when only epidermis was used (reporting "epiderm") = 0; when dermatomed skin was used (reporting "derm") = 300 default value; when full thickness skin was used (reporting "full thickness" or "scissors") = 900.
- **Wind velocity.** Since all predictions were made assuming occlusive conditions (immobile vehicle) the wind velocity was set to the default indoor scenario (0.165 m/s, Figure 2.6).

c) Observations during calculations

- **Batch run.** Data were entered for 48 compounds under different solvent conditions and concentrations and for various skin types and durations of exposure, resulting in overall 1586 entries. It was only possible to perform a batch run for up to 200 entries where the same skin type (dermis, epidermis, full thickness, scissors) and duration of exposure (6h, 8h, 10h, 24h) was selected. The data were grouped accordingly.
- **Solvent/water partition coefficient $K_{s/v}$.** It was not possible to run the model for organic solvent-based concentrates of substances for which the solvent/water partition coefficient $K_{s/v}$ was not specified. In these cases, predictions were only made for dilutions of these formulations, where water was assumed to be the appropriate solvent, as previously explained.
- **The specific case of Imidazole/BAS 590 02 F (Prochloraz-manganese):** This was identified as a polymeric complex with a MW of 1632 g/mol, for which it was not possible to obtain a SMILES code and retrieve a CAS number in EPI Suite. It was considered reasonable to

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

assume that the polymer would be too large to be the absorbed species. Instead, it appears plausible that prochloraz monomer is released and becomes (in part) dermally absorbed. Thus, the released monomer (Prochloraz) was considered the relevant species of the active substance. It is noted that velocity and extent of release of prochloraz monomer is not known, however, and this needs to be taken into account when assessing predictivity of the model against the EFSA dataset.

2.7.1.8 Model 8 – Potts & Guy 1995

Model Description

In addition to their classical model for predicting skin permeability (Potts & Guy, 1992), Potts and Guy later developed a partitioning-diffusion equation, based on free-energy relationships (Potts & Guy, 1995). Following this approach, partitioning is related to molecular volume (MV), and hydrogen bond donor and acceptor activities (H_d, H_a, respectively). After consideration and rearrangement of formulas associated with the free-energy relationships approach and the assumption that transport through lipid membranes is achieved by free volume fluctuations within the lipid domain, the following (simplified) relationship was derived:

$$\log k_p = a_1 * MV - a_2 * H_d - a_3 H_a + a_4$$

Equation 2.7.18

Multiple regression analysis on Equation 2.7.18 with aqueous permeability coefficients of 37 non-electrolytes yielded the following formula:

$$\log k_p \left[\frac{cm}{s}\right] = 10^2 * 2.56 * MV - 1.72 H_d - 3.93 H_a - 4.85$$

Equation 2.7.19

Where MV is the molecular volume and is given in $^{cm^3}/_{mol}$ ($1\text{Å}^3 = 10^{-24} cm^3$), H_d is the sum of hydrogen donor activity (acidity), and $\mathrm{H_a}$ is the sum of hydrogen acceptor activity (basicity) (Abraham *et al.*, 2002).

Model Implementation

Equation 2.7.19 was applied to calculate the permeability coefficient k_p. Hydrogen bond donor and acceptor activities (acidity/basicity) were calculated as described in section 2.6.2.

In the derived acidity/basicity values it was noticed that some non-feasible negative values were predicted by the model. One solution to that could be to set all negative values equal to 0 and continue with the calculation of $\log k_p$ with these "corrected" values. However, it was decided to keep the non-feasible acidity and basicity values as predicted by the model, by assuming that maybe negative basicity values would be derived by the model's attempt to overcompensate for very high acidity values and vice versa. For the sake of good modelling application practices this assumption needs to be further elucidated. Similarly to the other models, the calculated k_p was converted to cm/h by multiplying it with 3600 and then translated to %DA as described in section 2.5.1 by combination with the DAME model yielding two different %DA values, since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

2.7.2 Mixture Models

2.7.2.1 Model 9 – Riviere Brooks 2007 No. 1

Model Description

In order to develop algorithms that calculate permeability coefficients after exposure to mixtures, Riviere and Brooks (2007) elaborated three existing single compound based LFER models (Potts & Guy (1992), Hostynek & Magee (1997), Abraham & Martins (2004)) by taking into consideration mixture-related contributions. This was done by incorporation of an additional parameter, defined as Mixture Factor (MF), that describes the physical-chemical properties of the mixture based on those of its components. All three LFER models were tested by linear regression in order to identify the best fitting MF from a series of available physical-chemical properties (e.g. molecular volume, hydrogen bonding, TPSA, pKa). The dataset used for the derivation of the mixture models is composed of 288 data points/treatment combinations (12 different compounds in 24 mixtures) obtained from flow-through porcine skin diffusion cells. This process yielded two well-fitting models, based on the Potts and Guy algorithm[15].

In the first model (model 9), the mixture factor (MF) is related to the topological polar surface area of the mixture, mTpSA, which is derived as described in chapter 2.7.2.4.

The algorithm for model 9 is defined as follows:

$$\log k_p \left[\frac{cm}{h}\right] = -2.05 - 0.04 * mTpSA - 0.3 * \log \mathrm{K_{ow}} - 0.0008 * MW, r^2 = 0.69$$

Equation 2.7.20

Model Implementation

Equation 2.7.20 was implemented in Excel, with mTpSA expressed in Å² as calculated from TpSA values of the mixture components, $\log \mathrm{K_{ow}}$ corresponding to the octanol/water partition coefficient of the a.s., and MW corresponding to the molecular weight of the a.s.. Calculation of the mixture factor (mTpSA) is described in section 2.7.2.4. The model was applied only to experiments conducted with products for which detailed information on the composition could be retrieved. For the methodology followed for the full product characterization please refer to section 2.6.3. Similarly to the other models, the calculated k_p was translated to %DA as described in section 2.5.1 by combination with the DAME model yielding two different %DA values, since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

2.7.2.2 Model 10 – Riviere Brooks 2007 No. 2

Model Description

The model was derived by the methodology described above (2.7.2.1). In this second model (model 10), the mixture factor (MF) is related to the hydrogen bond acceptor count of the mixture, mHA, which is derived as described in chapter 2.7.2.4.

The algorithm for model 10 is defined as follows:

$$\log k_p \left[\frac{cm}{h}\right] = -1.12 - 1.19 * mHA - 0.3 * \log \mathrm{K_{ow}} - 0.00081 * MW, r^2 = 0.67$$

Equation 2.7.21

[15] The rest of the model fittings discussed in the publication (associated with Hostynek & Magee, Abraham & Martins models) were not examined in the present study and are thus not introduced here.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Model Implementation

Equation 2.7.21 was implemented in Excel, with mHA calculated from HA values of the mixture components, $\log K_{ow}$ corresponding to the octanol/water partition coefficient of the a.s. and MW corresponding to the molecular weight of the a.s.. Calculation of the mixture factor is described in section 2.7.2.4. The model was applied only to experiments conducted with products for which detailed information on the composition could be retrieved. For the methodology followed for the full product characterization please refer to section 2.6.3. Similarly to the other models, the calculated k_p was translated to %DA as described in section 2.5.1 by combination with the DAME model yielding two different %DA values, since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

2.7.2.3 Model 11 – Atobe 2015

Model Description

The feedforward artificial neural network (ANN) model developed by Atobe *et al.* (2015) predicts the permeability coefficient k_p of permeants by additional consideration of the physical-chemical properties of the vehicle. Input parameters for the training of the model were the octanol water partition coefficients ($LogP_{ow}$) and the molecular weights (MW) of 359 samples for which literature data on $logk_p$ were available. The training dataset comprised hydrocarbons, alcohols, aldehydes, ketones, ethers, esters, carboxylic acids, amines and amides, which were dissolved in various pure and mixed solvents. The ANN consisted of one input layer (input parameters: MW of permeant, $LogP_{ow}$ of permeant, $LogP_{ow}$ of vehicle), five hidden layers, and one output layer for the response variable k_p. The reported values of MW and $LogP_{ow}$ for the substance and the vehicle were in the ranges of 18.02 to 765.05, -3.70 to 7.88 and -1.39 to 8.72, respectively.

The resulting ANN model was compared to its multiple linear regression model (MLR) by calculating the root mean square errors (RMSE). The RMSE of 0.68 for the ANN model was lower than the one of the MLR model (0.89), indicating better performance of the ANN model (Atobe *et al.* 2015). For the 10-fold cross validation, an RMSE of 0.72 was calculated for the ANN model.

Model Implementation

The model was reconstructed on the basis of the training set published in the paper. The reconstruction was realized in MATLAB 2016b using Neural Network Toolbox, with 3 inputs, 20 hidden layers, and one output. The inputs corresponded to variables MW and $LogP_{ow}$ of the permeant, and $LogP_{ow}$ of the vehicle (also $mLogP_{ow}$), as provided in the paper, while the output corresponds to the variable $Logk_p$ (with k_p in cm/h). The RMSE after 10-fold cross validation was 0.68 for the reconstructed model with 20 hidden layers, while the RMSE of the original model was 0.72 after cross-validation according to Atobe *et al.* (2015).

The reconstructed ANN model was further used to estimate the value of $Logk_p$ given the input values for MW and $LogP_{ow}$ of the active substance, and $mLogP_{ow}$ of the mixture for 154 sets of data from the EFSA dataset. The parameter $mLogP_{ow}$ is defined as a Mixture Factor (MF) describing physical chemical properties of the vehicle mixture and is calculated from $LogP_{ow}$ of the individual mixture components as described in section 2.7.2.4.

Similarly to the other models, the calculated k_p was translated to %DA as described in section 2.5.1 by combination with the DAME model yielding two different %DA values, since two different t_{lag} scenarios were considered (t_{lag}=0; t_{lag}= calculated).

2.7.2.4 Calculation of Mixture Factors (MF)

The mixture models examined here describe individual compound penetration by additionally considering the effects of the co-formulants/vehicle, expressed as Mixture Factor (MF). Therefore, the results obtained by these models do not reflect the absorption of each single component of the

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

product, but only of the main a.s.. The mixture factors were calculated by summing up the mathematical product of the weight percentage and the respective parameter value for each of the identified components of a mixture. This was done for each of the three needed parameters. This process can be described using the following equations:

Total weight:

$$W_{tot} = W_{a.s.} + W_{c.f.1} + \ldots + W_{c.f.n}$$

Relative contribution of each component to the mixture:

$$RC_{a.s.} = \frac{W_{a.s.}}{W_{tot}};\ RC_{c.f.1.} = \frac{W_{c.f.1}}{W_{tot}};\ldots;\ RC_{c.f.n.} = \frac{W_{c.f.n}}{W_{tot}}$$

$$RC_{mixture} = \sum RC_{a.s.} + RC_{c.f.1} + \cdots + RC_{c.f.n} = 1$$

Relative contribution of the mixture and diluting agent (water) to the experimentally tested item:

$$RC_{tested} = RC_{mixture} * \frac{C_{tested}}{C_{concentrate}}$$

$$RC_{water} = 1 - RC_{tested}$$

Calculation of the mParameter (mP)

$$mP = \sum RC_{a.s.} * P_{a.s.} + RC_{c.f.1} * P_{c.f.2} + \cdots + RC_{c.f.n} * P_{tested} + RC_{water} * P_{water}$$

with $c.f._1$ - $c.f._n$ corresponding to co-formulants 1 to n.

For water, TpSA was 1Å^2 and HA was taken as 1. LogPow was assumed as -0.50 for water. For TpSA, HA and LogP_{ow} of a.s. and co-formulants, values retrieved from pubchem were preferred when available, otherwise computed values with Molinspiration (TpSA and HA) or EPI Suite (LogP_{ow}) were used (see also 2.6.3).

For weight, the concentration of the a.s. in the product [g/kg or g/l] and the co-formulant content [g/kg or g/l] were taken. $C_{concentrate}$ corresponds to the experimental concentration of the a.s. when tested as concentrate, and was taken for the whole experiment including the respective dilutions. For the studies where no tested concentrate was provided in the database, the concentrate concentration ($C_{concentrate}$) values were taken from the composition of the product for which also the co-formulant data were collected.

2.8 Statistical Analysis

2.8.1 Variability between replicates and the adjusted logit scale

As demonstrated in Appendix B of the EFSA (2017) guidance on dermal absorption and further confirmed in Appendix H of the present document, an "adjusted logit scale" is statistically appropriate when analyzing and presenting dermal absorption data. The adjusted logit scale was used in many places when averaging replicates from a block: the adjusted logit was computed for each replicate, the sample mean was computed using the adjusted logit values and the result was restored to the absorption scale by undoing the adjusted logit transformation.

Most of the plots which appear in the Results use the adjusted logit scale on one or both axes. Adjusted logit scales are used to enhance the communication and extraction of information. Some additional description and justification is provided in Appendix H.

2.8.2 Correlation analysis

2.8.2.1 Rank correlation analysis

Spearman rank correlation was used for comparison of predicted versus reported data. Analyses were performed on individual data and on block aggregates representing averages between replicates of one experiment. Analysis was also performed on coarse aggregates representing averages of all data for one active substance in order to check the possibility that correlations might be stronger for aggregates representing larger numbers of data.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Correlations were calculated using the whole data-set and also separately for concentrates and dilutions and the 4 formulation groups used in EFSA (2017) when considering the setting of default values: solids, water-based, organic solvent based and other. Averaging was performed on untransformed or logit-transformed data. Kendall's tau and P-values derived from Kendall's tau were calculated to assess the quality of the correlation, supported by visual inspection of respective plots.

2.8.2.2 Analysing the influence of covariates by restricting correlation analysis to subsets

A method for selecting subsets to evaluate was required which could be applied systematically with reasonable computational effort and which would allow a subset to be determined by more than one covariate. It was also important that the resulting subsets would have natural structure and would represent contiguous ranges of values for covariates

The building block is to divide the range of each covariate into 4 sub-ranges with roughly equal numbers of data in each sub-range. Exactly equal numbers are often not possible and the sub-ranges used were the best that could be found given the often discrete nature of the distributions of covariates. For a single covariate, the possible subsets are then made from contiguous groups of sub-ranges. Subsets involving multiple covariates are defined as intersections of subsets defined by single covariates.

It was decided to evaluate all subsets which are larger than some minimum size threshold which was set at 10% of the total number of cases for each covariate analysis in order to constrain the computational effort to be manageable. For the same reason, a limit was also set on the total number of covariates. Careful coding was required in order to avoid excessive computation time.

A method was needed to measure the extent to which a subset had a correlation which was worth considering. The basic principle was to take into account both the magnitude of correlation and the associate P-value for the test of no correlation. Due to skewness, outliers and non-linearity, non-parametric correlation methods were used; for calculating P-values, the test based on Kendall's tau was used because P-value computations were found to be more stable than using Spearman's rank correlation. However, for reporting correlations, Spearman's rank correlation was used due its greater familiarity and ease of interpretation as the usual Pearson product-moment correlation applied to ranks of data. The informal rules applied to identify interesting subsets were: (i) to require a P-value at least as significant as for all cases in covariate analysis; and (ii) to require a rank correlation at least as large as for all cases. The P-value requirement was relaxed for the analysis of model 7 due to a lack of interesting subsets using the original criteria. This approach involves comparison of the relative size of P-values. The P-values need to be interpreted with care due to the fact that: (i) many sub-sets are being evaluated; and (ii) the random sampling assumptions of Kendall's test are invalid for a dataset with hierarchical structure (active substances, formulations, etc).

2.8.2.3 Selection of covariates for correlation analysis

Eleven descriptors where selected for the purpose of covariate/correlation analysis:

- It is well established that molecular weight (MW) and octanol-water partition coefficient ($LogP_{ow}$) impact dermal absorption. Both parameters are included in various prediction models. In database found as: "Molecular Weight [g/Mol]" and "Log PO/W".
- The EFSA Guidance on dermal absorption (2012) and (2017) recommends selection of default or worst-case assumptions in the absence of measured data based on concentration of the active substance (EFSA, 2012) or the type of formulation and whether a concentrate or a dilution thereof is assessed (EFSA, 2017). Accordingly, parameters on concentration ("Conc"), concentrate/dilution status ("Concentrate") and formulation ("MergedForm") were included in the covariate analysis. The parameters are to be found in the database as: "Concentration Tested (g/L or g/kg)" for "Conc", and "Type of concentration tested (concentrate, dilution 1-3)" for "Concentrate". The data on the formulation ("MergedForm") were derived using

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

information on the formulation type (in database as "Formulation type") and assignment of formulation groups as defined in table B.2 of the EFSA Guidance (EFSA, 2017), see also Table 2.5 of the present report.

- Water solubility (S_{aq}) can limit the availability of the substance for absorption from water based formulations. Therefore, when measured information was available, S_{aq} (in the dataset named as "Water solubility [mg/L at pH 7 and 20/25 °C] (EFSA conclusion, or (d)RR / DAR / RAR / RAC opinion)") was included in the analysis. Otherwise, computed S_{aq} were considered (in the dataset named as "Water solubility computed (IJC) pH = 7 [mg/L]").
- Partitioning into and out of the SC is, among others, pH dependent and should thus relate to the ionization status of the permeant. For that purpose pKa was also considered for covariate analysis. Experimental (or computed, whenever experimental values not available) pKa (in dataset named as"pKa at 20 °C (EFSA conclusion, or (d)RR / DAR / RAR / RAC opinion)" or "pKa computed (IJC)") were used for the analysis.
- Another parameter considered in order to assess the influence of charge, was the sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which were hydrogen-bond acceptors, QO.N (in dataset named as "QO.N [e] (IJC)".
- Similarly, hydrogen bond formation may limit diffusivity, since permeants can bond to skin structures and therefore not reach the systemic circulation. To explore this, the sum of hydrogen bond donor and acceptor numbers (HB) was considered as covariate (= the sum of "HD [no] (pubchem)" and "HA [no] (pubchem)", as found in the dataset).
- The difference between the energies of the highest and the lowest occupied molecular orbital are frequently used to describe reactivity of the molecule. Therefore, the difference between "ELUMO [eV] (MOPAC2016, AM1 method)" (as found in the dataset) and "EHOMO [eV] (MOPAC2016, AM1 method)" (as found in the database) was also explored as covariate.
- Finally, melting point was considered as covariate, since it is imput parameter in two of the models (model 6 and 7) and is also influencing water solubility and thus absorption. Measured values (in the database named as "MP [°C] (EFSA conclusion, or (d)RR / DAR / RAR / RAC opinion)") were preferred. When unavailable, predicted values (in the database named as "MP [°C] (Episuite)") were used.

The table below lists potential covariates that were not included in the analysis and provides a brief justification.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 2.7: Substance properties not included in covariate analysis and justification for non-inclusion.

Parameter	Parameter name in dataset	Justification
S_{oct}: solubility in octanol	"solubility in octanol [g/L at 20/25 °C] (EFSA conclusion, or (d)RR / DAR / RAR / RAC opinion)"	Parameter is linked to octanol-water partition coefficient ($LogP_{ow}$) and water solubility (S_{aq}) which were both already included (see above).
H_d/H_a Hydrogen bond donor/acceptor activity	"Hydrogen bond donor activity = hydrogen bond acidity" OR "Hydrogen bond acceptor activity = hydrogen bond basicity"	A correlation with the hydrogen bond donor/acceptor counts (HD/HA) is expected. The latter was preferred because it is available from public database. HD/HA were included in the covariate analysis as sum of no. of hydrogen bond donors and acceptors (see above).
TpSA: Topological polar surface area	"TpSA [Å²] (pubchem)" OR "TpSA [A²] (molinspiration)"	Parameter is linked to octanol-water partition coefficient ($LogP_{ow}$) and molecular weight (MW) which were both already included (see above).
MV or Ve: Molecular volume or Van der Waals effective volume	"MV [A³] (molinspiration)" OR "Ve, van der Waals effective volume [A³] (IJC)"	Some correlation with molecular weight (MW) which was already included (see above).
computed HD/HA: hydrogen bond donor/acceptor counts	"HD [no.] (molinspiration)" OR "HA [no.] (molinspiration)"	For HD and HA number, information from public sources was preferred and available (already included, see above).
experimental or computed BP: boiling point	"BP [°C] (EFSA conclusion, or (d)RR / DAR / RAR / RAC opinion)" OR "BP [°C] (Episuite)"	Typically correlated to/predicted from further structural properties. Very little experimental (reported) data. Mostly computed values of unknown reliability.
experimental or computed VP: vapour pressure	"VP [Pa at 25°C] (EFSA conclusion, or (d)RR / DAR / RAR / RAC opinion)" OR "VP [mmHg at 25°C] (Episuite)"	As for boiling point (BP), VP is typically correlated to/predicted from BP and, to a lesser degree, further structural properties. In addition, experimental conditions restrict/control loss of material through evaporation and experiments with low recovery were not included in the EFSA database
Qh: sum of the net atomic charges of the hydrogen atoms bound to nitrogen or oxygen atoms	"Qh [e] (IJC)"	QO.N. (the sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which were hydrogen- bond acceptors) was included instead (see above).
MR: molar refractivity	"MR [10^6 m³/mol] (IJC)"	Computed values of unknown reliability.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

2.8.3 Bayesian Modelling

The Bayesian random effects (BREM) statistical modelling approach used to analyse the predictive performance of models evaluated in this report is similar to that taken in EFSA (2017). The main differences are:

i. that EFSA (2017) did not have the component corresponding to the model prediction and had instead more structure relating to dilutions versus concentrations and formulation groups than the model selected for use in section 3.5.2.3;
ii. modelling of variation between replicates. EFSA (2017) assumed a homogeneous normal distribution for variation between replicates. Based on further data analysis, a refined "heavy-tailed" t-distribution model was used in this report while the homogeneity assumption was retained.

Both EFSA (2017) and the modelling presented here use normal distributions for the random effects components which are treated as additive on adjusted-logit scale. The Bayesian model is described further in Appendix I.

2.8.4 Software

All calculations were carried out using R (R Core Team, 2017) with the support of a number of additional packages: Auguie (2017) and Wickham (2009, 2017) to assist with preparing graphics; Højsgaard & Halekoh (2016), Komsta & Novometsky (2015) and Wickham (2011) for general data analysis; Plate (2016) and Wickham & Bryan (2017) for data management; Plummer (2016), Plummer et al (2006) and Su and Yajima (2015) for Bayesian modelling and computation.

2.9 Survey on regulatory needs and questions in the context of dermal absorption modelling

A "Questionnaire on the practical applicability and potential regulatory implementation of *in silico* tools for the prediction of pesticides dermal absorption" was prepared considering also feedback by EFSA and the European Crop Protection Association (ECPA) and is provided as Supporting Information (file name: "Consultation Report and Questionnaire.pdf"). The purpose of this questionnaire was to receive feedback from Member State Competent Authorities (MSCAs) on regulatory needs and raise technical questions that should be considered and addressed in the statistical review of model predictions.

The questionnaire consisted of totally twelve (12) questions organised in three sets with increasing level of specificity and technical detail. The first set of questions concerned general issues on the use of *in silico* tools:

1. Please describe your involvement in the risk assessment of pesticides.
2. Do you currently use any *in silico* tools for regulatory purposes, such as prediction of toxicity?
3. Have you had any training on the use of *in silico* tools for regulatory purposes?

The second set of questions was related to the use of *in silico* tools for dermal absorption:

4. Are you aware of the *in silico* tools currently available for the prediction of dermal absorption of pesticides (active substance)?
5. Are you aware of the *in silico* tools currently available for the prediction of dermal absorption of pesticide mixtures (products)?
6. Do you currently use *in silico* tools for the prediction of dermal absorption of pesticides for regulatory purposes?
7. Do you consider that the use of *in silico* tools for the prediction of pesticide dermal absorption is linked to high level of uncertainty?
8. Would you use *in silico* approaches only as screening tools that may not be considered in replacement of experimental data on dermal absorption?

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

9. Would you consider *in silico* approaches as valuable tools that could be used in replacement of experimental data on dermal absorption?

Other more specific technical questions were addressed at the end of the questionnaire:

10. Which of the following input parameters do you consider critical and applicable for an *in silico* tool to be used for the prediction of dermal absorption?
    - Formulation type (EC, SC, WP, WG, etc).
    - Exact formulation composition
    - Main solvent(s) (in case of liquid formulations)
    - Particle size distribution (in case of solid formulations)
    - Content of the active substance in the formulation
    - Octanol-water partition coefficient (log POW)
    - Molecular weight of active substance
    - Permeation constants
    - Other physicochemical parameters
    - Other
11. Do you consider that a probabilistic or a deterministic *in silico* model would be more suitable for regulatory implementation?
12. Do you consider that classification (e.g. product classification) is relevant for an *in silico* tool to be used for the prediction of dermal absorption?

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3 Assessment/Results

3.1 Regulatory needs and questions in the context of dermal absorption modelling

Thirteen (13) responses on the "Questionnaire on the practical applicability and potential regulatory implementation of *in silico* tools for the prediction of pesticides dermal absorption" were received from ten MSCAs, i.e. Finland (FI), Austria (AT), the Czech Republic (CZ), Germany (DE), Denmark (DK), Spain (ES), France (FR), the Netherlands (NL), Poland (PL) and the United Kingdom (UK). AT, ES and PL provided two separate responses from two different experts. Since the answers from the individual experts from AT, ES and PL were not identical, they were considered as separate expert opinions rather than national positions. So, the results are presented from all 13 completed questionnaires.

All experts responded that they were involved in the mammalian toxicology evaluation and risk assessment of active substances and their products, confirming their relevance to the topic.

In general, the use of *in silico* tools is driven by the regulatory needs as described in current legislation and related guidelines. Since *in silico* tools are currently not recommended for the prediction of dermal absorption of pesticides (EFSA, 2017), none of the experts is using them, routinely. Furthermore, there is a general scepticism on the use of *in silico* tools for dermal absorption of pesticides for several reasons. Firstly, very few of the experts have the technical knowledge on how to use the available tools reliably, highlighting the need for training on the topic. In addition, the multifactorial nature of dermal absorption is recognised as a parameter limiting the *in silico* tools reliability.

However, several experts would be willing to use *in silico* tools for the prediction of pesticide dermal absorption if validated methods are available and are recommended by regulatory bodies (e.g. EFSA). In this case training will be necessary.

Based on the feedback from MSCA, the following questions could be considered as starting points in the selection of models from the literature for statistical review of model predictivity or in the statistical analysis itself:

- Is the applicability domain of the model relevant for the prediction of dermal absorption of pesticides?
- Is the model only restricted to aqueous solutions or neat test substances or is it also applicable to other formulation types (e.g. organic solvent-based, solid formulations)?
- Is the model applicable for the prediction of dermal absorption of active substances in complex mixtures?
- Does the tested model predict dermal absorption with a clearly defined level of uncertainty? Is it possible to lower the estimated uncertainty for specific formulation types?
- Does the model take into consideration the complexity of agrochemical formulations, including the physicochemical properties of the formulations (e.g. surface tension, pH, viscosity) and their in-use dilutions, the active substance (e.g. physical state, molecular weight, lipid/water partition coefficient Ionization, water solubility, pKa, local skin effects), the vehicle(s) and other co-formulants (e.g. solubility, volatility, distribution in *stratum corneum*, excipients effect on the *stratum corneum* pH; 4)?
- Does the model take into consideration the broad variety of experimental conditions (e.g. state of occlusion (occlusive, semi-occlusive or non-occlusive) or preparation of the skin sample (epidermal sheet, dermatomed skin or full-thickness skin), skin area in contact with vehicle and duration of exposure)?
- Does the model take into consideration differences in skin parameters due to species (e.g. rat or human), anatomical site, temperature (of skin), hydration of *stratum corneum*, damage to *stratum corneum*, metabolism, diseased skin, desquamation blood and lymph flow?

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

- Does the model take into consideration product classification with regard to skin irritation/corrosion?
- Is the model easily accessible to the expert? Is the model validated?
- Is it possible to repeat the prediction with the same level of uncertainty considering specific instructions by the model developer or in case training is provided?

A detailed analysis of the completed questionnaires is provided as Supporting Information (file name: "Consultation Report and Questionnaire.pdf") to this report. The feedback received is used as guiding information for the analysis of dermal absorption models in this study.

3.2 Literature Search and Appraisal

The literature search performed as described in chapter 2.1 revealed a total of 2212 records published between 1990 and June 2017 in the public domain. Assessment of relevance of these records based on title and abstract, if available, identified 288 publications describing one or more mathematical models for *in silico* prediction of dermal absorption. Based on the first stage of the scientific review, 103 of the 288 papers originally regarded as relevant were excluded from further consideration as they proved to not be relevant when the full publication was assessed. The remaining 185 publications were subdivided over four categories based on type of publication (review/original study) and on type of model (single substance or mixture model) as shown in Table 3.1.

Table 3.1: Categorisation of relevant papers. Between brackets the final numbers after the scientific review: two papers attributed to "single substance models" were misclassified and are on mixture models, while one paper on single substance models had been classified as "original" while it was a review paper.

Type of compounds	Type of paper		Total
	Original	Review	
Single substances	128 (125)	41 (42)	**169 (167)**
Mixtures	15 (17)	1	**17 (18)**
Total	**143 (142)**	**42 (43)**	***185***

Only the publications describing original work were entered for the second stage of the scientific review as the reviews were not expected to contribute additions to the pool of models extracted from the original studies.

The 142 original papers described a total of 233 models, 188 predicting absorption of single substances (in the overwhelming majority from an aqueous vehicle) and 45 predicting absorption of substances from mixtures. A detailed list is provided as Supporting Information (file name: "Model Review.xlsx").

Most models described were algorithm based: 141 out of 188 single substance models and 37 out of 45 mixture models. 122 single substance models and 31 mixture models passed the criterion of having a clear algorithm and being suitable for reconstruction. In addition, machine learning tools were identified of which 15 single substance models and 7 mixture models in principle can be reconstructed as the training sets with the descriptor values are available. The suitability of the models for pesticide risk assessment was examined taking into account the outcome on the survey described in chapter 3.1. Models deemed not suitable for pesticides were excluded, leaving 84 single substance models and 36 mixture models. The algorithmic models that proved to be possibly relevant for the prediction of dermal absorption in the framework of pesticide risk assessment are further detailed in Appendix D. Restriction to the most recent versions of all models leaves 75 single substance models and 36 mixture models (Table 3.2).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.2: Overview of the selection of models from the 142 relevant original papers for ranking.

Selection criterion	Single substances		Mixtures		Total
	Algorithm based	Machine Learning	Algorithm based	Machine Learning	
Reconstructible	122	15	31	7	**175**
Suitability pesticide RA	70	14	29	7	**120**
Most recent version	**61**	**14**	**29**	**7**	***111***

The overwhelming majority of the single substance models that are, in principle, suitable for predicting pesticide dermal absorption, predict the permeation constant k_p, some predict the maximum flux J_{max} and only one predicts percentage absorption (see Table 3.3). The same accounts for mixture models, where 27 models predict k_p, while the remainder predict the steady state flux Jss as absorption parameter.

Table 3.3: Overview of absorption parameters predicted by the selected models. Cells report number of models predicting a specific absorption parameter.

Model type	Single substances			Mixtures		Total
	% abs.	J_{max}	k_p	J_{ss}	k_p	
Algorithm based	0	6	55	3	26	**90**
Machine learning	1	0	13	6	1	**21**
Total	**1**	**6**	**68**	**9**	**27**	***111***

Models identified were based on a large number of physicochemical descriptors as input parameters for prediction. A comprehensive list of these descriptors and terminologies is provided in Appendix D, Table D.3.

Many of the algorithm based single substance models use simple physicochemical properties as input parameters, of which 16 employ only MW and/or P_{ow} (see Table 3.4). Single substance machine learning models tend to use more complex descriptors, as do mixture models in general.

Table 3.4: Overview of descriptor types used by the selected models. Cells report number of models using a specific descriptor type. Descriptor types are sorted from simple types to more complex types. Models using a more complex descriptor type may also employ in addition less complex types of descriptors, e.g. models using complex physicochemical properties may also need P_{ow} as an input. Algo. Stands for Algorithm based models and ML for Machine learning models.

Most "complex" descriptor type used	Single subst. mod.		Mix. mod.	
	Algo.	ML	Algo.	ML
Simple physicochemical properties: MW	1	0	0	0
Simple physicochemical properties: P_{OW}	3	0	0	0
Simple physicochemical properties: P_{OW} and MW	11	1	1	1
Simple physicochemical properties: other	14	1	3	0
Abraham	7	1	13	0
Complex physicochemical properties	1	2	0	0
Complex structural and/or quantum-chemical descriptors	24	9	12	6
Total	**61**	**14**	**29**	**7**

Complex descriptors may be difficult to obtain for all chemicals because of the availability of the software tools needed or because they have to be generated in a specific chemical test when not

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

predictable from chemical structure. This may be especially problematic for mixture models, as for each prediction of absorption of an active substance also the descriptors of other constituents of the mixture need to be available.

The full results of the first and the second stage of this evaluation are provided in Appendix D and as Supporting Information (file name: "Model Review.xlsx").

3.3 Model Implementation

For the 111 models representing most recent versions identified as described in chapter 3.2, the availability of required input data as a pre-requisite for implementation was assessed. A weighted procedure was chosen, resulting in a score of 0 when one or more of the essential input parameters was considered inaccessible and giving a maximum score of 5 when all input data is provided in the EFSA dataset (refer to chapter 2.4 for details). In Table 3.5 are presented the different levels of availability with their respective frequencies:

Table 3.5: Frequency of parameter availability scores for models identified in the systematic literature search (rounded to two significant figures).

Parameter availability		Frequency	
Score	**Description**	**No. (total 284)**	**%**
5	provided for a.s. in EFSA dataset	10	3.5
4	provided for a.s. in DAR or study report	16	5.6
3	data or QSPR tool open access	37	13
2	QSPR proprietary but available in consortium	37	13
1	replacement by surrogate feasible	76	27
0	not accessible with reasonable effort	108	38

In the second step of model selection, all individual parameter availability scores were integrated into one single model parameter score (MPS) using the following equation:

$$MPS = \sqrt[n]{PS_1 * PS_2 * PS_3 * \ldots * PS_n}$$

Equation 3.3.1

where PS_x corresponds to the parameter scores of each parameter of a given model

Models with many input parameters were neither preferred nor discriminated. Models with MPS < 1 were excluded from the selection process. As a result, 43 single substance models and 7 mixture models received an overall model parameter availability score above zero. These were taken forward to the next stage.

These 50 comprised models were

- based on artificial neural networks (ANN) and, more conventionally, mathematical algorithms,
- using as input either few physicochemical descriptors as MW and LogP, a more complex combination of many physico-chemical values or quantum-chemical descriptors,
- predicting the permeability constant k_p, maximum flux J_{max} or percentage dermal absorption %DA.

In accordance with the project objective, models were grouped accordingly and representative approaches were selected to ensure wide coverage for single substance as well as mixture models. A reserve list of models was created to serve as backup list in case ANN based models cannot be reconstructed or unforeseen problems are encountered during model implementation, e.g. with collection of individual input parameters.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In addition, since the classical Potts & Guy (1992) algorithm is the most common applied model in literature, it was decided to also include this one as a "historic reference" despite the existance of more recent version with minor revisions.

An overview of selected models for implementation within the project is provided in Table 3.5.

Following collection of required input data as described in chapters 2.6.2 and 2.6.3, model implementation was performed as outlined in chapter 2.7. All selected models could be successfully implemented and the reserve list was not used. Notably, interpretation of ambiguous instructions to the complex MS-Excel based Model 7 initially caused implausible predictions which required corrections to the approach the supplied tool was used.

Results of all predictions are provided asSupporting Information (file name: "Human *in vitro* PPP EFSA dataset with add parameters and predictions.xlsx") and were subjected to detailed statistical analysis. All detailed calculation steps are available upon request at EFSA.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.6: Overview of selected single substance models and mixture models for model implementation.

Paper	Model Parameters	Model Algorithm	Training Dataset
Single substance models (SSM)			
Paper#19 Potts R. O. and Guy R. H. 1992	**MW:** molecular weight **P_{ow}:** octanol-water partition coefficient	log k_p [cm/h] = 0.71 LogP_{ow} – 0.0061 MW – 2.7	data compiled by Tayar *et al.* 1991 and Flynn 1990
Paper#13 Magnusson B. M. and Anissimov Y. G. and Cross S. E. and Roberts M. S. 2004	**MW:** molecular weight	$\log J_{max}\left[\frac{mol}{cm^2*h}\right]$ = –0.0141MW – 4.52	alcohols, steroids, phenol derivatives, and other chemical classes (Magnusson dataset)
Paper#19 Potts R. O. and Guy R. H. 1995	**H_a:** Hydrogen bond acceptor activity (also B - solute overall (summation) hydrogen bond basicity) **H_d:** Hydrogen bond donor activity (also A - solute overall (summation) hydrogen bond acidity) **MV:** molecular volume	log k_p [cm/h] = 0.0256MV – 1.72HD – 3.93HA – 1.29	alcohols, steroids, phenol derivatives, and other chemical groups
Paper#44 Dancik Y. and Miller M. A. and Jaworska J. and Kasting G. B. 2013	**P_{ow}:** octanol-water partition coefficient **MW:** molecular weight **VP:** vapour pressure **MP:** melting point **BP:** boiling point **#double bonds** **#triple bonds** **#aromatic and non-aromatic rings** **optional:** **ρ:** Molecule density **Saq:** Water solubility **measured k_p:** permeation coefficient **pKa:** negative logarithm (to the base 10) of the dissociation const	n/a: complex model predicting %DA	n.a.
Paper#49 Frasch H. F. 2002	**MW:** molecular weight **P_{ow}:** octanol-water partition coefficient	k_p [cm/h] = k_sc · k_aq/k_sc+k_aq k_aq = 0.1151 cm/h k_sc = K_mv x D /l* log K_mv = 0.59 log K_ow - 0.024	alcohols, steroids, phenol derivatives, and other chemical groups (Flynn, 1990 database)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Paper	Model Parameters	Model Algorithm	Training Dataset
		log D = log _D_cor_D_lip/1+EXP- log K_cor_lip + 0.1974 - 0.3668 log D_cor/D_lip/0.2488 + -0.134 log D_cor/D_lip log D_cor/D_lip = -0.0087 MW l* = 0.003 x 1 - 0.9113 log K_cor/lip + 0.9896 log _Kcor_lip^2 + 0.3111 log K_cor/lip^3 log K_cor/lip = -0.8075 log K_ow + 2.4194	
Paper#51 (1) Fu X. C. and Ma X. W. and Liang W. Q. 2002	**MV:** molecular volume **Qh:** the sum of the net atomic charges of the hydrogen atoms bound to nitrogen or oxygen atoms **QO.N:** the sum of the absolute values of the net atomic charges of oxygene and nitrogen atoms which were hydrogen-bond acceptors **EHOMO:** the energy of the highest occupied molecular orbital (eV) **ELUMO:** the energy of the lowest unoccupied molecular orbital (eV)	log k_p [cm/h] = 3.69 MV - 2.86 Qh - 2.19 QO.N - 0.033 EHOMO - 0.22 ELUMO - 1.20	alcohols, phenol derivates, carboxylic acids, steroids (Abraham *et al.* (1997) data set)
Paper#51 (2) Fu X. C. and Ma X. W. and Liang W. Q. 2002	**MV, Qh, QO.N, EHOMO, ELUMO**	n/a: ANN model	alcohols, phenol derivates, carboxylic acids, steroids (Abraham *et al.* (1997) data set)
Paper#85 Milewski M. and Stinchcomb A. L. 2012	**MW:** molecular weight **P_{ow}:** octanol-water partition coefficient **MP:** melting point	log J_max $\left[\frac{nmol}{cm^2*h}\right]$= 4.6- 0.219 $LogP_{ow}$ - 0.0086 MW - 0.0102 (MP – 25)	alcohols, steroids, phenol derivatives, and other chemical classes (Magnusson dataset)
Single substance models (SSM) – Reserve list			
Paper#5 Buchwald P. and Bodor N. 2001	**N:** hydrogen bonding-related parameter for special functional groups	log k_p = 0.0208 log V_e – 0.723N – 2.69	steroids, alcohols, phenol derivatives, pharmacologically active molecules,

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Paper	Model Parameters	Model Algorithm	Training Dataset
(alternative to Paper#51)	of molecules **Ve:** Van der Waals effective molecular volumes		and other chemical groups
Paper#112 Thomas J. and Majumdar S. and Wasdo S. and Majumdar A. and Sloan K. B. 2007 (alternative to Paper#13/#85)	**log Soct:** solubility in octanol **logSaq:** solubility in water **MW:** molecular weight	log J_max = -2.574 + 0.5861 log Soct + 0.4139 log Saq - 0.00440 MW	phenols, hydrocortisone esters, aliphatic alcohols, aliphatic carboxylic acids, analgetics, steroids, miscellaneous (edited Flynn database)
Paper#74 Kilian D. and Lemmer H. J. R. and Gerber M. and Du Preez J. L. and Du Plessis J. 2016 (alternative to #49)	**MW:** molecular weight **P_{ow}:** octanol-water partition coefficient	log k_p = 0.739 log K_ow - 0.0089 MW - 2.36	published datasets from Lian *et al.* (2008), Moss and Cronin (2002), Wilschut *et al.* (1995)
Mixture models (MM)			
Paper#9 (1) Riviere J. E. and Brooks J. D. (2007)	**MW:** molecular weight **P_{ow}:** octanol-water partition coefficient **mTPSA:** mixture Topological Polar Surface Area difference	$logk_p$ [cm/h] = -0.04 mTPSA - 0.03 $LogP_{ow}$- 0.00080 MW - 2.05	substituted phenols, organophosphates, triazine herbicides
Paper#9 (2) Riviere J. E. and Brooks J. D. (2007)	**MW:** molecular weight **P_{ow}** : octanol-water partition coefficient **mHA:** mixture number of hydrogen bond acceptors	$logk_p$ [cm/h] = -1.19 mHA - 0.03 $logP_{ow}$ - 0.00081 MW - 1.12	substituted phenols, organophosphates, triazine herbicides
Paper#15 (2) Atobe T. and Mori M. and Yamashita F. and Hashida M. and Kouzuki H. (2015)	**MW(chemical):** molecular weight **P_{ow} (chemical):** octanol-water partition coefficient **P_{ow} (vehicle):** octanol-water partition coefficient	n/a: ANN using MW(chemical), P_{ow} (chemical), P_{ow} (vehicle)	hydrocarbons, alcohols, aldehydes, ketones, ethers, esters, carboxylic acids, amines, amides
Mixture models (MM) – Reserve list			
Paper#15 (1) Atobe T. and Mori M. and Yamashita F. and Hashida M. and Kouzuki H. (2015) (alternative to Paper#9 (2))	**MW(chemical):** molecular weight **P_{ow} (chemical):** octanol-water partition coefficient **P_{ow} (vehicle):** octanol-water partition	$logk_p$ [cm/h] = −0.193 x $logP_{ow}$ (chemical) x $logP_{ow}$ (vehicle) + 0.00124 x MW(chemical) x $logP_{ow}$ (vehicle) - 0.00476 x MW(chemical) + 0.0184 x	hydrocarbons, alcohols, aldehydes, ketones, ethers, esters, carboxylic acids, amines, amides

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Paper	Model Parameters	Model Algorithm	Training Dataset
	coefficient	$logP_{ow}$ (vehicle))^2 - 0.00000352 x MW(chemical)^2 - 2.23	
Paper#9 (3) Riviere J. E. and Brooks J. D. (2007) (alternative to Paper#9 (1))	**MR:** molar refractivity **HBA:** hydrogen bond acceptors = counts of hydrogen bond donors **HBD:** hydrogen bond acceptors = counts of hydrogen bond donors **mTPSA:** mixture Topological Polar Surface Area difference	$logk_p$ [cm/h] = -0.04 mTPSA - 0.48 MR + 0.09 HBA - 0.42 HBD - 0.49	substituted phenols, organophosphates, triazine herbicides

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.4 Statistical analysis – general aspects

3.4.1 The influence of lag-time

As described in the model implementation sections (2.7) the predicted permeability constant k_p or maximum flux J_{max} derived by selected models were combined with DAME to yield %DA values. Two types of predictions were made for each model, one based on a lag time set to 0 ("worst case"), and one with a calculated lag time. This applies to all selected models with the exception of model 7. An initial analysis relating the results of the two sub-versions, showed that the differences between the two sub-versions of each model (tlag=0 vs calculated tlag) are not sufficient to merit producing a separate in-depth evaluation for each sub-version. When applicable, the calculated tlag sub-versions are used hereafter for all models in further assessment. The results are presented in detail in Appendix J.

3.4.2 Variability between replicates and averaging

The plots provided in the next chapters and appendices show averages of replicates within blocks of replicates ("block aggregates"). In principle, this is done to reduce the risk of bias due to varying numbers of replicates and one would also expect some reduction in statistical noise in relationships due to averaging. This should also be reflected in the outcome of the following correlation analyses.

One might expect that predictions for a set of replicate measurements would all be the same and would therefore not need averaging. However, the models use the "Applied Concentration (µg/cm^2) PER CELL" in calculations and this is not always constant for a group of replicates even though the "Concentration Tested (g/L or g/kg)" is constant within each group. When available, the computed values were derived by using as "applied concentration" the one that was introduced per replicate (Franz cell), i.e. in that particular cell. However, this was not always reported/measured at the individual replicate level. In these cases, the introduced concentration means from many replicates/cells of the same experimental block were used. For this reason, the relationship between predictions for pairs of replicates from within the same block of replicates was examined, as a measure of how closely related the single replicate values are to their averaged ones. As an example of the actual, the following plot shows the relationship between predictions by model 1 for pairs of replicates from within the same block of replicates. This should be noted as a limitation to the analysis of the influence on between replicate variability. In addition, it provides an additional reason to perform correlation analysis preferably on block aggregates in order to reduce the risk of bias due to varying numbers of replicates.

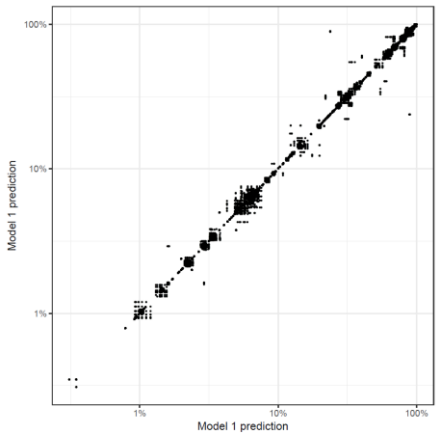


Figure 3.1: Relationship between %DA predictions by model 1 for pairs of replicates from within the same block of replicates. X- and Y-axes correspond to predicted values for different replicates of the same block.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.5 Performance of Single Substance Models

3.5.1 Models 1, 2, 3, 4, 5, 6, 8

3.5.1.1 Correlations

Plots depicting the relationship between the experimental % Dermal Absorption (%DA) values and predicted values do not reveal any correlation. This is true both for individual replicates and for block aggregates (i.e. averages of measured replicates within blocks of replicates and corresponding average predictions). In addition to linear scales, log and logit transformations/scaling and combinations thereof were employed as an attempt to extract any visual relation, however with no success. Rank plots also do not show any correlation; this is not surprising since rank plots are known for reducing apparent association.

Below, an example of the relationship between averages of measured replicates within blocks of replicates and corresponding average predictions with model 1 is provided. The calculated lag time version of the model was used (for explanations refer to 2.5.1 & Appendix J). Similar correlations are observed for the other models (models 2, 3, 4, 5, 6, 8) and are to be found in Appendix K.

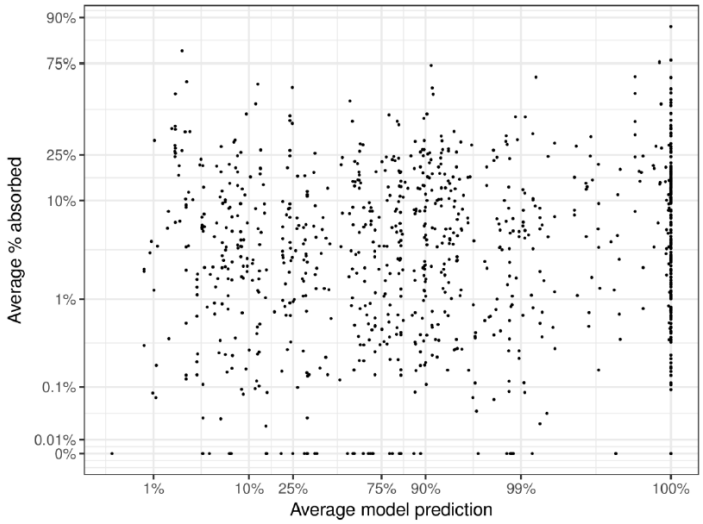


Figure 3.2: Relationship between the experimental % Dermal Absorption (%DA) values and the predicted ones with model 1 (M1), the combination of Frasch (2002) and Buist (2010) (DAME) model. Adjusted logit scales (Appendix H) are used for both axes, each data point corresponds to averaged values for a block of replicates.

The visual observations stated above were confirmed by Spearman rho and Kendall tau rank correlations between measured and predicted %DA.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.7: Rank correlations between measured and predicted %DA for each model (or model version). The second and third columns refer to the correlations for individual replicates and corresponding %DA predictions. The fourth and fifth columns refer to the correlations for block aggregates, i.e. the mean of replicates within blocks and the corresponding averaged %DA prediction. Averages were computed using the adjusted logit scale (see section 2.8.1). For explanations on the models refer to Appendix G.

Model	Spearman rho - replicates	Kendall tau - replicates	No. of replicates	Spearman rho - block aggregates	Kendall tau - block aggregates	No. of blocks
1	0.11	0.07	6232	0.11	0.07	947
2	0.12	0.08	6232	0.12	0.08	947
3 (exp. Saq)	0.01	0.00	6050	0.00	0.00	919
3 (calc. Saq)	-0.01	-0.01	6232	-0.02	-0.01	947
4	0.07	0.05	6224	0.07	0.05	946
5	0.07	0.05	6200	0.08	0.06	943
6 (exp. Saq)	0.04	0.03	6050	0.03	0.02	919
6 (calc. Saq)	0.07	0.05	6232	0.07	0.05	947
6 (exp. Saq, nT)	0.01	0.01	6050	-0.01	-0.01	919
6 (calc. Saq, nT)	-0.03	-0.02	6232	-0.04	-0.03	947
8	-0.03	-0.02	6232	-0.03	-0.02	947

3.5.1.2 Influence of covariates

The relationships described above were broken down into subsets[16] /plots that additionally distinguishbetween:

i. dilution status (concentrate/in-use dilution)
ii. dilution status (concentrate/in-use dilution) and formulation types (organic solvent/ solid/water based/other)

Visual inspection of the relationship plots (Appendix L, Figure L.1 and Figure L.2) and rank correlation coefficients (Table 3.8a&b) revealed no apparent improvement for any of the subsets justifying statistical follow-up.

Table 3.8a: Spearman rank correlations between measured and predicted %DA for each model (or model version), separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used. For explanations on the models refer to Appendix G.

Model	Dilutions	Concen-trates	Solids	Water based	Organic solvent
1	0.18 (n=595)	0.17 (n=352)	0.14 (n=107)	0.03 (n=391)	0.16 (n=406)
2	0.20 (n=595)	0.18 (n=352)	0.20 (n=107)	0.04 (n=391)	0.16 (n=406)
3 (exp. Saq)	0.01 (n=576)	0.09 (n=343)	-0.13 (n=105)	-0.01 (n=374)	0.01 (n=399)
3 (calc. Saq)	-0.01 (n=595)	0.07 (n=352)	-0.14 (n=107)	-0.03 (n=391)	0.00 (n=406)
4	0.14 (n=594)	0.10 (n=352)	0.19 (n=106)	-0.04 (n=391)	0.18 (n=406)
5	0.14 (n=592)	0.06 (n=351)	0.15 (n=106)	-0.01 (n=388)	0.16 (n=406)
6 (exp. Saq)	0.05 (n=576)	0.15 (n=343)	-0.12 (n=105)	0.00 (n=374)	0.04 (n=399)
6 (calc. Saq)	0.12 (n=595)	0.14 (n=352)	-0.1 (n=107)	0.01 (n=391)	0.10 (n=406)
6 (exp. Saq, nT)	0.00 (n=576)	0.02 (n=343)	0.06 (n=105)	-0.03 (n=374)	-0.03 (n=399)
6 (calc. Saq, nT)	-0.02 (n=595)	-0.02 (n=352)	-0.1 (n=107)	-0.03 (n=391)	0.00 (n=406)
8	-0.06 (n=595)	-0.00 (n=352)	-0.13 (n=107)	-0.05 (n=391)	-0.09 (n=406)

[16] Sub-categorisation is in line with the proposals outlined in the revised EFSA Dermal Absorption guidance (2017)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.8b: Kendall tau rank correlations between measured and predicted %DA for each model (or model version), separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used. For explanations on the models refer to Appendix G.

Model	Dilutions	Concentrates	Solids	Water based	Organic solvent
1	0.12 (n=595)	0.12 (n=352)	0.09 (n=107)	0.02 (n=391)	0.11 (n=406)
2	0.13 (n=595)	0.12 (n=352)	0.13 (n=107)	0.03 (n=391)	0.11 (n=406)
3 (exp. Saq)	0.01 (n=576)	0.06 (n=343)	-0.09 (n=105)	-0.01 (n=374)	0.01 (n=399)
3 (calc. Saq)	0.00 (n=595)	0.05 (n=352)	-0.09 (n=106)	-0.02 (n=391)	0.00 (n=406)
4	0.09 (n=594)	0.07 (n=352)	0.13 (n=106)	-0.02 (n=391)	0.13 (n=406)
5	0.09 (n=592)	0.04 (n=351)	0.10 (n=107)	-0.01 (n=388)	0.11 (n=406)
6 (exp. Saq)	0.04 (n=576)	0.11 (n=343)	-0.08 (n=105)	0.00 (n=374)	0.03 (n=399)
6 (calc. Saq)	0.09 (n=595)	0.10 (n=352)	-0.06 (n=107)	0.01 (n=391)	0.07 (n=406)
6 (exp. Saq, nT)	0.00 (n=576)	0.02 (n=343)	0.04 (n=105)	-0.02 (n=374)	-0.02 (n=399)
6 (calc. Saq, nT)	-0.01 (n=595)	-0.01 (n=352)	-0.06 (n=107)	-0.02 (n=391)	0.00 (n=406)
8	-0.04 (n=595)	0.00 (n=352)	-0.09 (n=107)	-0.04 (n=391)	-0.06 (n=406)

3.5.1.3 Predictive capacity

In addition to the above exploration of the predictive capacity of the models by considering plots of measured absorption versus predicted absorption for each model and by calculating rank correlations between measured and predicted absorption, a more formal, empirical analysis of the predictive capacity of model 1 was performed.

It needs to be noted that empirical analysis has two weaknesses. Firstly, the true absorption is not available and has to be approximated by measured absorption which may be either at the level of an individual cell or be the average of a block of replicates. Secondly, empirical analysis is specific to the particular substances, formulations as well as test conditions considered and does not attach any uncertainty to the results when extrapolating to consider a new active substance or formulation.

For the analysis for model 1, averages of blocks of replicates were used. Even though, as stated above, no substantial correlation between measured and predicted absorption is evident, there is still a pattern of under- and over-prediction. Predictions which are low will tend to under-predict and predictions which are high will tend to over-predict. This is shown in Table 3.9 and Figure 3.3 (left panel) for model 1, where ratios of measured to predicted absorption are broken down by the range of the prediction. However, this pattern does not have any real predictive meaning. There would be no advantage of using the model predictions for a particular case than simply picking a random value from all the predictions made by the model (Figure 3.3, right panel). The apparently striking correlation in Figure 3.3 (left panel) arises from the fact that predicted absorption is included both as the variable on the horizontal axis and as the denominator in the ratio used for the vertical axis. It is well known that such correlations are often spurious (Kronmal, 1993). Similar considerations apply to the rest of the models (models 2-6 & 8).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.9: Table for model 1 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor.

	Model absorption prediction (depending on range of predicted value)			
Number of blocks:	7	145	734	61
Model prediction is:	**0.1-1%**	**1-10%**	**10-100%**	**100%**
100 times lower than measured	0%	0%	0%	0%
20 times lower than measured	0%	11%	0%	0%
5 times lower than measured	0%	16%	0%	0%
2 times lower than measured	71%	32%	1%	0%
higher than measured	29%	53%	98%	100%
2 times higher than measured	29%	43%	94%	97%
5 times higher than measured	14%	21%	78%	89%
10 times higher than measured	14%	17%	60%	70%
100 times higher than measured	0%	1%	22%	26%

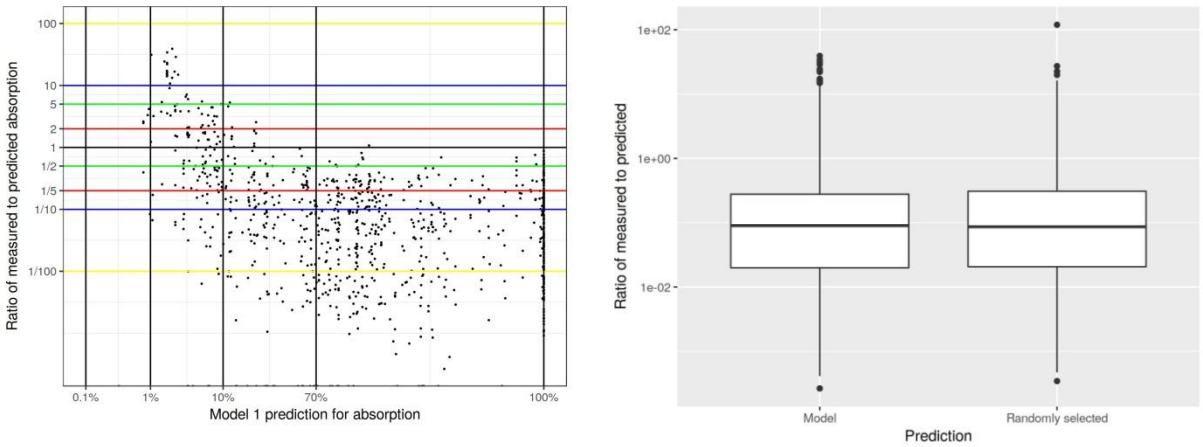


Figure 3.3: Left panel: ratio of measured to predicted absorption versus predicted absorption for model 1. Vertical and horizontal lines highlight the categories used in Table 3.9. Right panel: Box and whisker plots of ratio of measure to predicted absorption using either (i) the specific prediction from model 1 for the measurement or (ii) a randomly selected prediction from model 1. Averages of blocks of replicates were used for both cases.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.5.2 Model 7 – Dancik/NIOSH

3.5.2.1 Correlations

This part provides insight into the correlations between measured and model-predicted absorption values. Below, the relationship of the averages of replicates within blocks of measured replicates and their corresponding average model predictions is depicted. The use of averages aims to reduce the risk of bias due to varying numbers of replicates and the statistical noise in relationships due to averaging. Four versions of the same relationship, with different scales on the axes, are used. In each case, the measured absorption is on the vertical axis and the absorption predicted by model 7 is on the horizontal axis. The top-left panel shows the relationship using the unmodified percentage absorbed scale for both axes. The bottom-left panel uses the adjusted logit scale for the vertical axis. The bottom-right panel uses the adjusted logit scale for both axes. The top-right panel uses a rank scale for both axes. In other words, each measurement is replaced by its rank amongst all measurements and each prediction is replaced by its rank amongst all the predictions from the model. This panel does not depend on potentially arbitrary choices of scale and is the basis of the Spearman rank-correlations presented later.

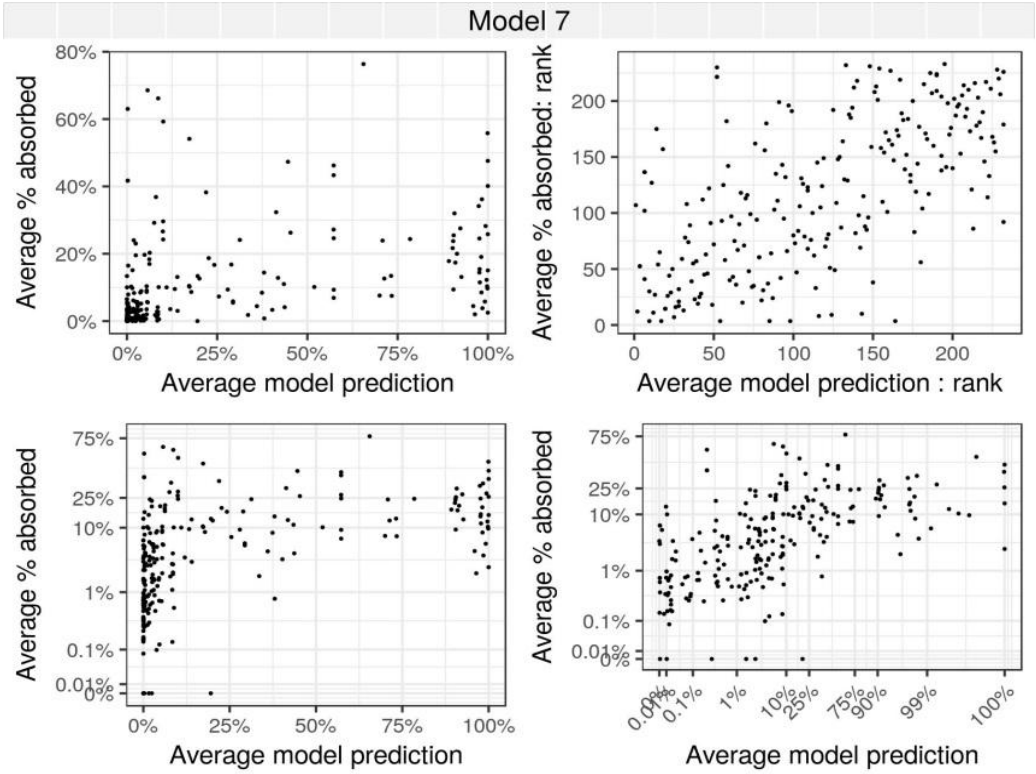


Figure 3.4: Relationship between the experimental % Dermal Absorption (%DA) values and the predicted ones with model 7 (M7), theNIOSH model. Top-left panel: unmodified (linear) percentage absorbed scale for both axes; bottom-left panel: adjusted logit scale for the vertical axis; bottom-right panel: adjusted logit scale for both axes; top-right panel: rank scale for both axes.

From Figure 3.4, it is visually apparent that model 7 exhibits a positive correlation between measured and predicted %DA values, as seen on the bottom-right (adjusted logit) and top-right (rank scale) plots. This is confirmed by Spearman and Kendall tau rank correlations (Table 3.10).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.10: Rank correlations between measured and predicted %DA for model 7. The second and third columns refer to the correlations for individual replicates and corresponding %DA predictions. The fourth and fifth columns refer to the correlations for block aggregates, i.e. the mean of replicates within blocks and the corresponding averaged %DA prediction. For averaging, logit-transformation was applied (see section 2.8.1).

Rank correlations	Spearman rho - replicates	Kendall tau - replicates	No. of replicates	Spearman rho - block aggregates	Kendall tau - block aggregates	No. of blocks
Model 7	0.61	0.43	1558	0.64	0.45	233

The correlations in the table suggest that model 7 is worthy of further statistical investigation. For subsequent analyses, use of block aggregates obtained by averaging at the logit-scale appears preferable and is used unless indicated otherwise.

3.5.2.2 Influence of covariates

The better overall rank correlations for model 7 warrant detailed analyses examining the influence of covariates on correlation. First, the impact of restriction to subsets defined by concentration status and merged formulation type on the correlation between measured values and those predicted by model 7 was evaluated (Figure 3.5 & Table 3.11). Concentration status was linked to both, measured values and those predicted by model 7, confirming the recommendations in the EFSA guidance for selection of default assumptions based on concentration status. Sub-categorisation into concentrates and in-use dilutions did, however, not improve the correlation between measured and predicted data (Figure 3.5, left panel). The correlations within the two categories are weak. There is evidence of predictive relationships within some sub-groups representing the different groups of formulation types (Figure 3.5, right panel & Table 3.11a&b) although the correlations are not very strong, and are never stronger than the correlation using all the data shown in Table 3.10.

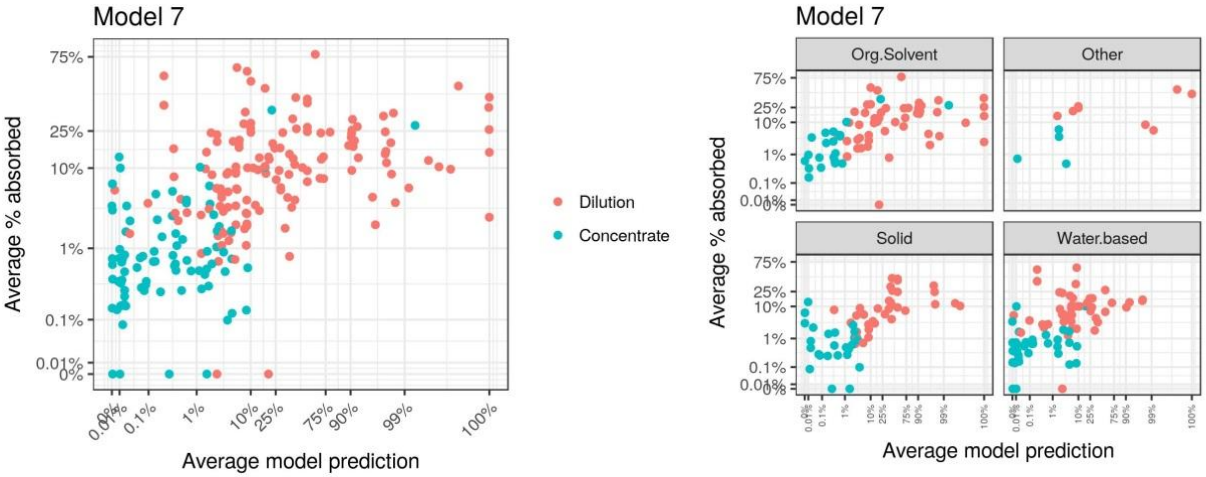


Figure 3.5: Relationships between the averaged experimental % Dermal Absorption (%DA) values and values predicted by model 7, divided (left) according to concentration status into concentrates and in-use dilutions; (right) by formulations groups and into concentrates and dilutions.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.11a: Spearman rho rank correlations between measured and predicted %DA for model 7, separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used.

Spearman Rank correlation	Dilutions	Concentrates	Solids	Water based	Organic solvent
Model 7	0.38 (n=145)	0.15 (n=88)	0.67 (n=61)	0.58 (n=84)	0.64 (n=74)

Table 3.11b: Kendall tau rank correlations between measured and predicted %DA for model 7, separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used.

Kendall tau Rank correlation	Dilutions	Concentrates	Solids	Water based	Organic solvent
Model 7	0.26 (n=145)	0.12 (n=88)	0.49 (n=61)	0.41 (n=84)	0.46 (n=74)

Other covariates

In addition to sub-setting for concentration status and merged formulation type, it was decided to perform additional detailed correlation analyses on more covariates as described in chapter 2.8.2.3. There are 233 data points (blocks of replicates) for which there is model 7 output. As described in 2.8.2.2, the system of making four categories for each covariate was used. The results of the subcategorization are shown in Appendix L, Table L.1. For all subsets defined by that system which had at least 24 data points (10% threshold), correlation analysis was performed.

No subset had a smaller P-value than the original data without sub-categorisation using Kendall's tau test for the correlation between model 7 output and measured absorption (data not shown). When looking at subsets which had a P-value no more than 100 times larger than for the original data, two subsets with improved correlation coefficients – active substances with pKa up to 9.4 or water solubility of at least 6.8 mg/L - could be identified for further consideration.

Table 3.12: Improved correlation between measured and predicted dermal absorption using Model 7 for selected subsets of data.

Subset	log_10 P-value from Kendall test	Spearman rho	Kendall tau	n
All data	-24.2	0.64	0.45	233
pKa ≤ 9.4	-23.9	0.72	0.52	176
Water solubility (Saq) ≥ 6.8	-22.6	0.71	0.50	177

Note: Very small P-values should not necessarily be interpreted as reflecting "extraordinarily significant" findings, as large datasets nearly always produce extreme statistical significance even when the predictive/scientific significance may not be much and the P-value assumes that the data points are a random sample (whereas here, there are hierarchical structures with multiple points per active substance).

The subset of data with pKa>9.4 for active substances had a noticeably weaker correlation (Figure 3.6). This subset would represent basic substances that predominantly carry cationic charges at neutral pH. Potentially, these may cause electrostatic interaction with polyanionic biopolymers of the skin, which might not be appropriately captured in model 7. When excluding predictions for active substances with pKa>9.4, the correlation of model 7 with measured data improved from 0.64 to 0.72.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

For water solubility of active substances (Saq), an increase in correlation coefficient when restricting the analysis to solubilities ≥6.8 mg/L from 0.64 to 0.71 is observed (Table 3.12)., the figure below shows that data points relating to solubility below 6.8 mg/L (red points) within the subset containing substances with high pKa (bottom right in the panel) are less structured.

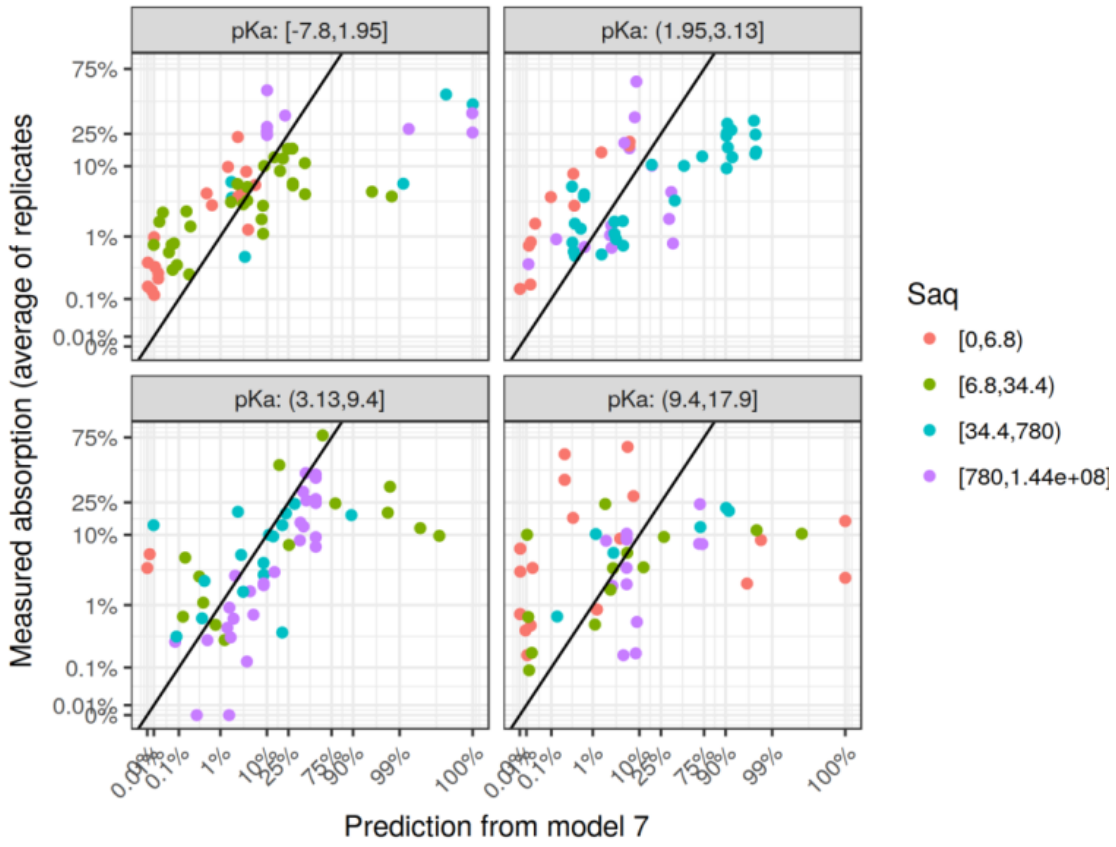


Figure 3.6: Relationship between the averaged experimental % Dermal Absorption (%DA) values and those predicted using model 7, subdivided into pKa and Saq (water solubility in mg/l) categories, as defined Appendix L, Table L.1.

Overall, Figure 3.6 illustrates that the P-values did not get smaller in the subsets because there is no really striking improvement in the pattern compared to the full dataset, neither when restricting pKa (compare individual figures of the panel) nor water solubility (compare colors). This cannot off-set the effect of the reduced number of data points on P-values.

3.5.2.3 Predictive capacity

In addition to the correlations analyses performed above (3.5.2.1 and 3.5.2.2), the predictive capacity of model 7 is analysed more formally by performing empirical analysis and statistical modelling. As noted in section 3.5.1.3, the empirical analysis performed in the present context faces some limitations. For that reason the predictivity was additionally explored by statistical/probabilistic modelling.

Here, the empirical analysis of the predictive capacity was performed with averages of blocks of replicates. Model 7 has some real predictive power as seen from the corresponding table and figure (Table 3.13 & Figure 3.7, left panel) and box and whisker plots which show that the "error" tends to be smaller using the case-specific prediction than using a random prediction (Figure 3.7, right panel).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.13: Table for model 7 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor.

	Model absorption prediction (depending on range of predicted value)				
Number of blocks:	28	31	77	79	3
Model prediction is:	**0.01-0.1%**	**0.1-1%**	**1-10%**	**10-100%**	**100%**
100 times lower than measured	7%	10%	0%	0%	0%
20 times lower than measured	**75%**	26%	3%	0%	0%
5 times lower than measured	**86%**	**55%**	10%	1%	0%
2 times lower than measured	**100%**	**74%**	29%	5%	0%
higher than measured	0%	10%	**53%**	**91%**	**100%**
2 times higher than measured	0%	3%	**39%**	**71%**	**100%**
5 times higher than measured	0%	3%	13%	**39%**	**67%**
10 times higher than measured	0%	3%	6%	14%	**33%**
100 times higher than measured	0%	3%	1%	0%	0%

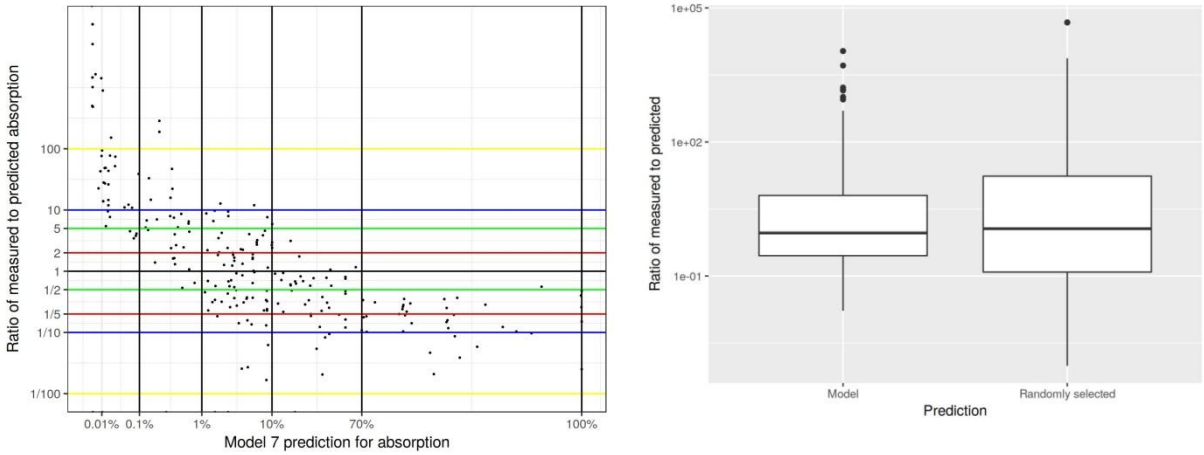


Figure 3.7: Left panel: Ratio of measured to predicted absorption versus predicted absorption for model 7. Vertical and horizontal lines highlight the categories used in Table 3.13. Right panel: Box and whisker plots of ratio of measured to predicted absorption using either (i) the specific prediction from model 7 for the measurement or (ii) a randomly selected prediction from model 7.

However, the scatter-plot of measured versus predicted absorption (Figure 3.4, lower-left panel) shows that model 7 still tends to over-estimate absorption when it makes a high prediction and to under-estimate absorption when it makes a low prediction. This can in principle be corrected by using a regression model as described below.

For the statistical modelling analysis, Bayesian modelling was employed. The aim was to develop a Bayesian random effects model (BREM) that can incorporate components relating to important features of the problem:

(A) the fact that, despite the correlation, model 7 exhibits biases in relation to low and high predicted absorption can be addressed by building a regression relationship between true absorption and model prediction. A key assumption made in doing so is that there is a real predictive relationship which is not, for example, due solely to differences between concentrates and dilutions.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

(B) the regression "error" term can be structured hierarchically:
- component for variation between active substances ("random effects" of active substance)
- component for variation between blocks of replicates for the same active substance ("random effects" representing different studies, concentrations, etc for measurements on the same substance)
- component for variation between replicates within a block

The box and whisker plots shown in Figure 3.8 demonstrate the homogeneity of within-block variability for dilutions and concentrates, across formulation groups and for different concentration levels. The quantile-quantile plot demonstrates that the variability between replicates is not normal, since there is clear evidence of heavy-tails (Figure 3.9).

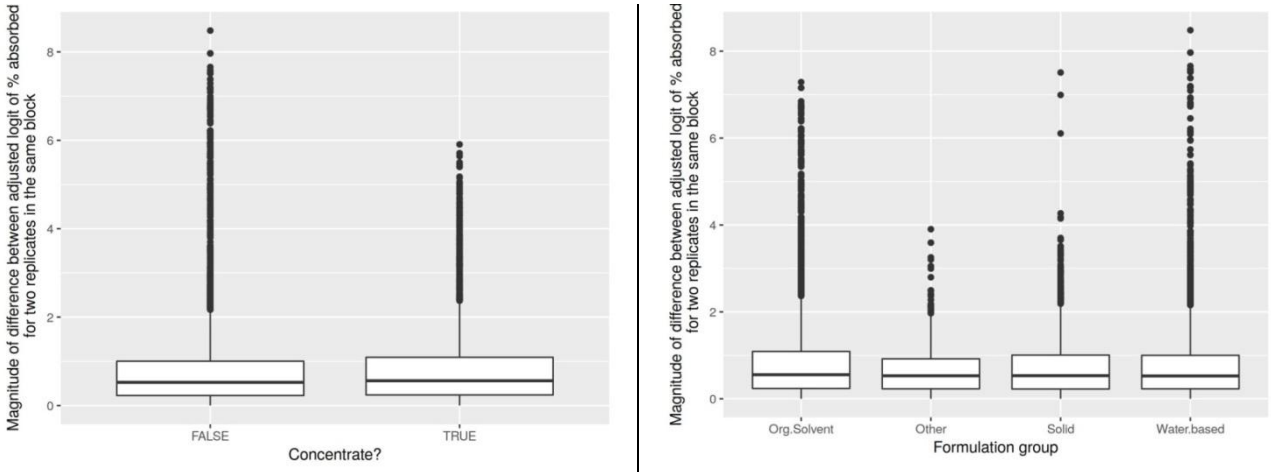


Figure 3.8: Boxplots assessing homogeneity of within-block variability for concentrates and dilutions (left) and for formulation groups (right).

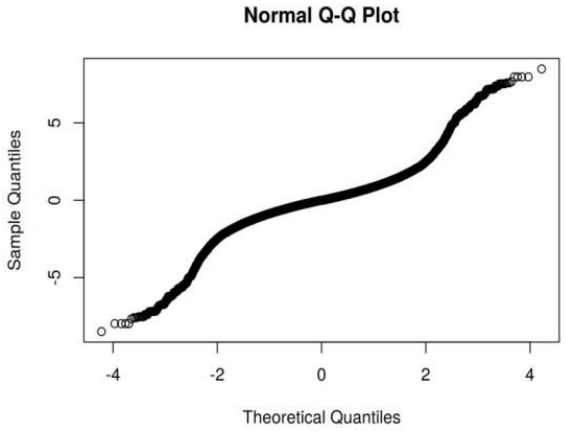


Figure 3.9: Quantile-quantile plot of differences of adjusted-logit values for pairs of replicates from the same block.

The regression relating true absorption to model prediction was based on a linear relationship between the adjusted-logit of true absorption and the logarithm (base 10) of the model prediction (as a percentage x%). This relationship is shown in the figure where the solid black curve shows where predicted and measured absorption are equal and the solid red line shows the line of best fit (least squares) corresponding to the regression carried out in the conventional way without hierarchical

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

structure for the regression errors. The logarithm was used because it results in a clearly linear relationship in Figure 3.10 whereas the bottom-right panel of Figure 3.4, which uses the adjusted-logit scale for the model prediction, has some non-linearity.

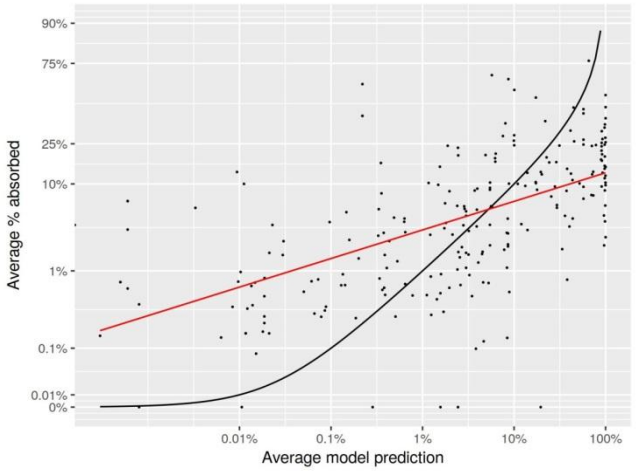


Figure 3.10: Measured absorption versus model 7 prediction using adjusted logit scale for the former and logarithm scale for the latter. The solid black curve is shows where measured equals predicted; it is not a straight line because the axes have different scales. The red line is the line of best fit (least squares) using the adjusted logit and logarithmic scales. i.e. it is the regression line for predicting the adjusted logit of measured absorption from the base 10 logarithm of the predicted absorption.

The observed rank correlation for the relationship in the figure is 0.64. The standard (Pearson product-moment) correlation for the strength of linear relationship in the regression is 0.59.

Table 3.14: Results of fitting the BREM for model 7.

BREM parameter	Estimate	95% credible interval	
		Lower limit	Upper limit
Intercept of linear relationship	-3.52	-3.87	-3.18
Slope of linear relationship	1.01	0.86	1.15
Standard deviation of variation between substances	0.98	0.71	1.33
Standard deviation of variation between blocks for the same substance	1.13	1.02	1.27
Scale of variation between replicates within a block	0.46	0.42	0.49
Degrees of freedom for within-block variation t-distribution	3.1	2.5	3.8

The BREM, described technically in 2.8.3 and Appendix I, has structure which allows evaluation of how well a model predicts true absorption and what kind of correlations should be expected from an empirical analysis of the type reported earlier. However, there is an ambiguity about which sources of variation should be excluded when defining "true absorption". Clearly, variability between replicates should be excluded. However, it is possible that some or all of the variation between studies should also be excluded. This is however more difficult to define and implement. Consequently, the focus here is on predicting the true absorption for each block of replicates.

Although the BREM includes a representation of the true absorption, the true absorption remains unknown. To assess the likely level of observed correlation between true absorption and measured absorption, two scenarios were considered: (i) where the model prediction is exactly correct; (ii)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

where the measured absorption is exactly correct. In both scenarios, the correlation was computed between the assumed true absorption and synthetic data with measurement variability added according to the BREM component describing variability between replicates. The process for generating synthetic data is random and so the correlation obtained varies if the process is repeated. 1000 synthetic data-sets were generated for each scenario. For scenario (i), 95% of the resulting rank correlations for individual cell data were in the range [0.969, 0.978] and and for averages were in the range [0.991, 0.996]. For scenario 2, the corresponding ranges were [0.923, 0.942] and [0.980, 0.990]. Clearly the observed correlations in Table 3.10 are much lower than this and this analysis indicates that they are not low simply due to variability of absorption measurements within blocks.

The BREM enables quantification of the probability that the model over- or under-estimates the true absorption by any specified amount at any specified level of model prediction. It also enables quantification of statistical uncertainty relating to the probability.

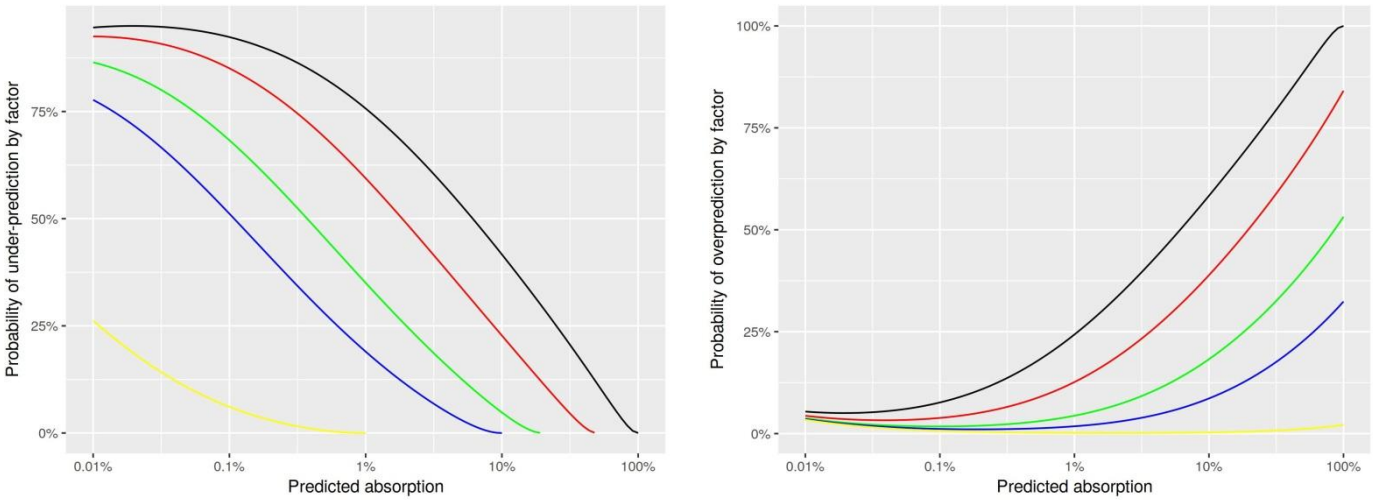


Figure 3.11: Probability that model 7 under-predicts (left) or over-predicts (right) true absorption by specified factor. Factors: 1 (black), 2 (red), 5 (green), 10 (blue), yellow (100).

Adjusted predictions of model 7

The BREM can also be used to adjust the prediction of model 7 in order to try to adjust for the tendency for predictions of low absorption to under-predict and for predictions of high absorption to over-predict. By passing the output of the model through the linear regression, it is possible to achieve better predictive performance. The method is to compute the (base 10) logarithm of the prediction from model 7, multiply by the estimate of the slope parameter from table 13 and add the estimate of the intercept parameter from table 13, and complete the prediction by inverting the adjusted-logit calculation.

For example, if the model predicts 0.5% absorption, the adjusted prediction can be obtained as follows. First apply the regression equation to find the adjusted-logit of the adjust prediction using the linear regression:

$$\ell = -3.52 + 1.01 \times \log_{10} 0.5 = -3.824$$

Then apply equations (2) and (3) from appendix H to undo the logit and then undo the shrinking used in the adjusted-logit:

$$f_{\text{shrunk}} = e^{-3.824}/(1 + e^{-3.824}) = 0.0214 \qquad f = 0.5 + \frac{0.0214 - 0.5}{0.9995} = 0.0211$$

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The adjusted prediction is 0.0211 = 2.11% absorption.

The estimated probability that the adjusted prediction exceeds the true absorption by a specified factor can then be calculated using the BREM. Figure 3.12 illustrates the resulting performance. It should be recognised that the adjustment and calculation of performance are both conditional on the BREM being a good statistical model for the data and on the data being appropriately representative.

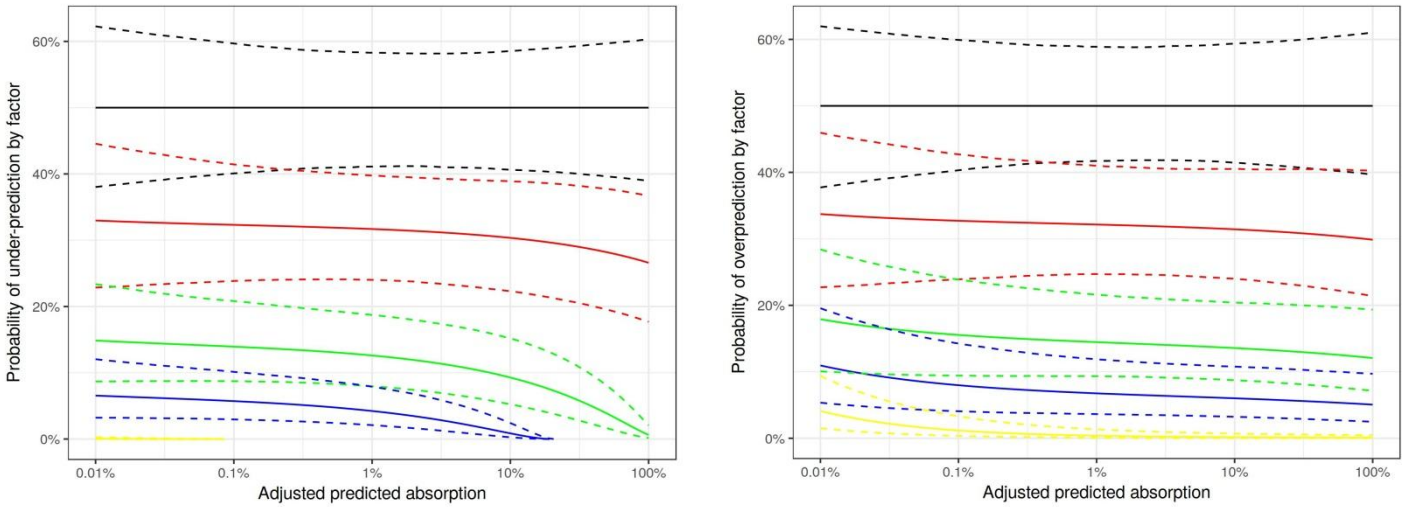


Figure 3.12: Estimated probability that adjusted model 7 under-predicts (left) or over-predicts (right) true absorption by a specified factor. Factors: 1(black), 2(red), 5 (green), 10 (blue), yellow (100). Solid lines represent the probability of over-/underestimating by a specified factor. The dashed lines indicate "uncertainty about the probability of over-/underestimating by the specified factor". To be more precise, for a fixed value on the horizontal axis (adjusted prediction from model 7), the dashed lines indicate a "95% credible interval for the frequency with which the adjusted prediction will over-/underpredict by at least the specified factor" where "frequency" refers to repeat use of the adjusted model for new substances. The uncertainty shown is the uncertainty from within the BREM and does not cover uncertainty about the structure of the statistical model used to relate measured absorption to predicted absorption.

It can be seen that the estimated probability of under- or over-estimating absorption by a specified factor is much less dependent on the level of absorption prediction when comparing the adjusted model to the original model. This indicates that that approach is reasonably successful. The probabilities of under-prediction fall away at high predicted absorption but this is a consequence of the fact that absorption cannot exceed 100%. For over-prediction, the estimated probabilities increase little for low predicted absorption and this is due to the adjustment made to the logit scale.

Despite the success in achieving reasonably homogeneous under- and over-prediction probabilities, there is still a high chance of substantial under- or over-prediction. This is due to the relatively weak correlation between predicted and measured absorption. The strength of correlation is a fundamental limitation on the process of adjusting model predictions.

Further considerations on uncertainties in model prediction are presented in chapter 3.8 focussing on regulatory relevance.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.6 Performance of Mixture Models

3.6.1 Models 9 and 10 – Riviere Brooks 2007 1&2

3.6.1.1 Correlations

Below, the relationship of the averages of replicates within blocks of measured replicates and their corresponding average model predictions is depicted. The use of averages aims to reduce the risk of bias due to varying numbers of replicates and the statistical noise in relationships due to averaging. For all these figures the calculated lag time version of the model was used (for explanations refer to 2.5.1 & Appendix J). Four versions of the same relationship, with different scales on the axes, are used. In each case, the measured absorption is on the vertical axis and the absorption predicted by models 9 and 10 is on the horizontal axis. The top-left panel shows the relationship using the unmodified percentage absorbed scale for both axes. The bottom-left panel uses the adjusted logit scale for the vertical axis only. The bottom-right panel uses the adjusted logit scale for both axes. The top-right panel uses a rank scale for both axes. In this representation, each measurement is replaced by its rank amongst all measurements and each prediction is replaced by its rank amongst all the predictions from the model. This panel does not depend on potentially arbitrary choices of scale and is the basis of the Spearman rank-correlations presented later.

Figure 3.13 reveals somewhat positive correlations between measured and predicted %DA for Models 9 and 10.These findings are in line with the Spearman rho and Kendall tau rank-correlations analysis presented in Table 3.15. The correlations in the table suggest that the mixture models 9 & 10 are worthy of further statistical investigation. For further analyses, use of block aggregates obtained by averaging at the logit-scale appears preferable and is used unless indicated otherwise.

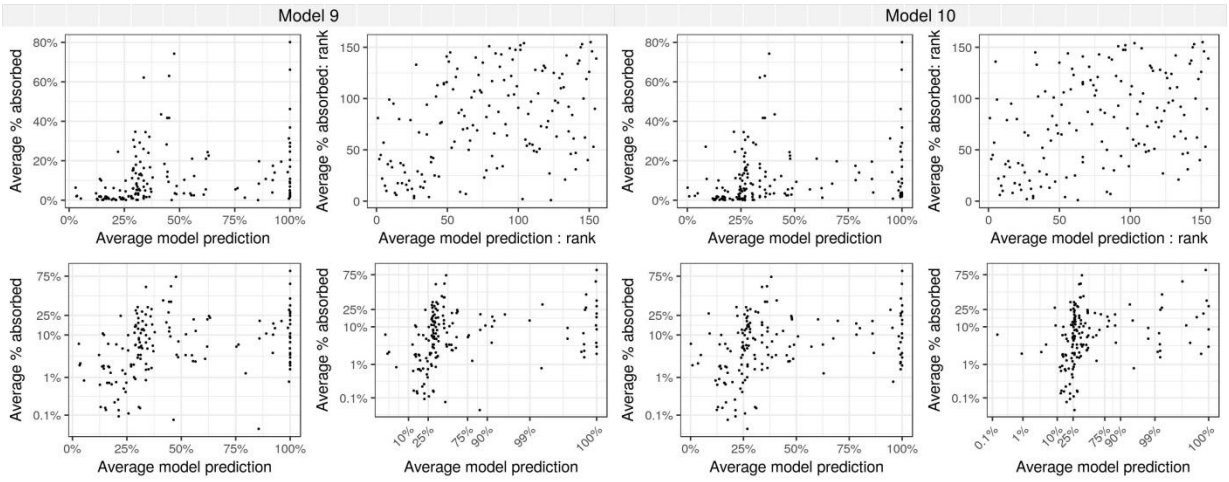


Figure 3.13: Relationship between the experimental % Dermal Absorption (%DA) values and the predicted ones with (left) model 9 (M9), the combination of Riviere & Brooks (2007) algorithm No.1 and Buist (2010) (DAME) model using calculated lag-time; and (right) model 10 (M10), the combination of Riviere & Brooks (2007) algorithm No.2 and Buist (2010) (DAME) model using calculated lag-time. Top-left panels: unmodified (linear) percentage absorbed scale for both axes; bottom-left panels: adjusted logit scale for the vertical axis; bottom-right panels: adjusted logit scale for both axes; top-right panels: rank scale for both axes.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.15: Rank correlations between measured and predicted %DA for models 9 and 10. The second and third columns refer to the correlations for individual replicates and corresponding %DA predictions. The fourth and fifth columns refer to the correlations for block aggregates, i.e. the mean of replicates within blocks and the corresponding averaged %DA prediction. Averages were computed using the adjusted logit scale (see section 2.8.1).

Model	Spearman rho Replicates	Kendall tau Replicates	Number of replicates	Block aggregates	Kendall tau Block aggregates	Number of blocks
9	0.39	0.26	945	0.43	0.29	155
10	0.37	0.25	945	0.41	0.27	155

3.6.1.2 Influence of covariates

The influence of covariates on the correlation detected in the previous section was examined by splitting the dataset into various subsets. First, the impact of restriction to subsets defined by concentration status and merged formulation type on the correlation between measured values and those predicted by mixture models 9 and 10 was evaluated (Figure 3.14 & Table 3.16).

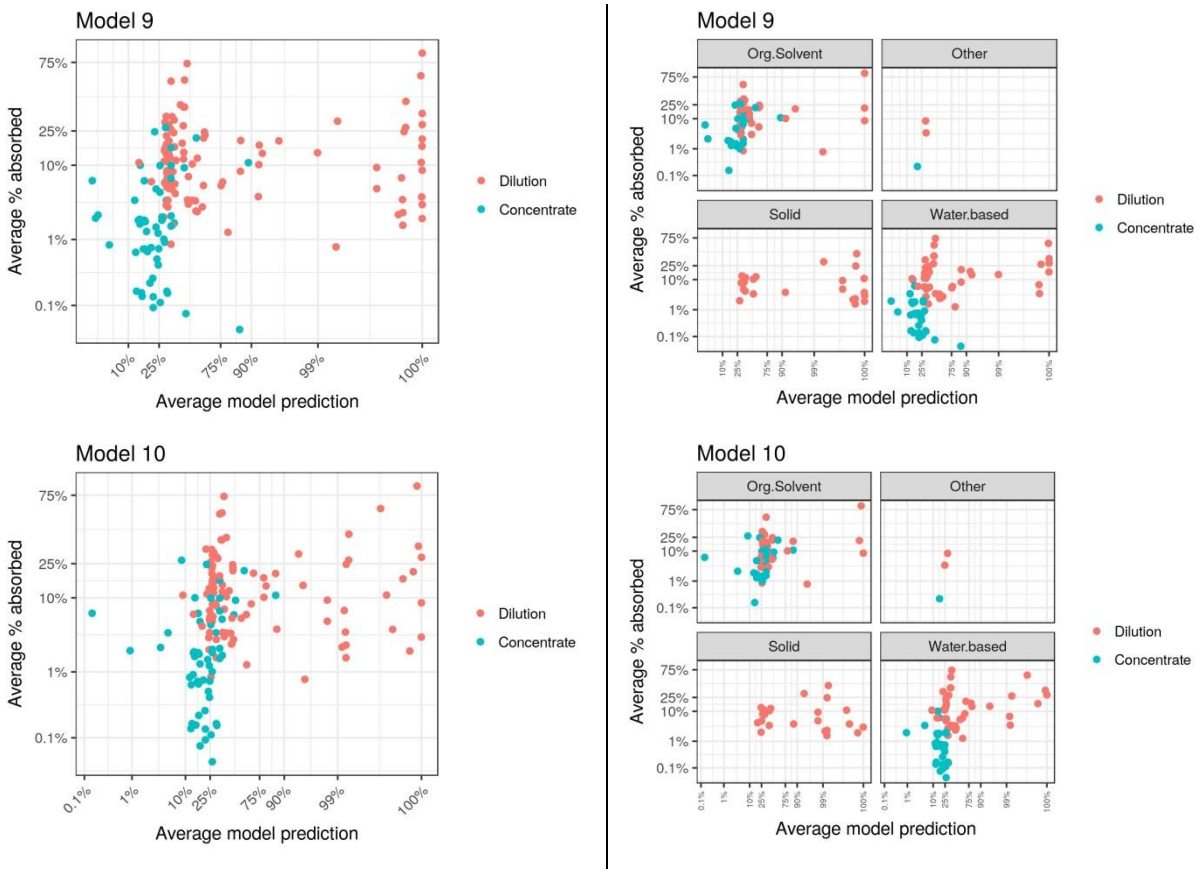


Figure 3.14: Relationships between the averaged experimental % Dermal Absorption (%DA) values and values predicted by models 9 (top) and 10 (bottom), divided according to (left) concentration status into concentrates and in-use dilutions; (right) by formulations groups and into concentrates and dilutions.

Concentration status was linked to both, measured values and those predicted by models 9 and 10, confirming the recommendations of EFSA for selection of default assumptions based on concentration status. Overall, subsetting for concentration status and formulation type did not lead to the

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

identification of stronger relationships. Only for water based formulations there appears to exist an improved relationship for both models 9 and 10; a finding also supported by the respective Spearman and Kendall tau rank correlations that are higher than the ones for the whole dataset.

Table 3.16a: Spearman rank correlations between measured and predicted %DA, separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used.

Model	Dilutions	Concentrates	Solids	Water based	Organic solvent
9	0.00 (n=99)	0.17 (n=56)	-0.14 (n=23)	0.51 (n=72)	0.41 (n=55)
10	0.03 (n=99)	0.20 (n=56)	-0.14 (n=23)	0.52 (n=72)	0.31 (n=55)

Table 3.16b: Kendall tau rank correlations between measured and predicted %DA, separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used.

Model	Dilutions	Concentrates	Solids	Water based	Organic solvent
9	0.00 (n=99)	0.10 (n=56)	-0.09 (n=23)	0.35 (n=72)	0.28 (n=55)
10	0.03 (n=99)	0.13 (n=56)	-0.08 (n=23)	0.35 (n=72)	0.21 (n=55)

Other covariates

In addition to sub-setting for concentration status and merged formulation type, additional correlation analyses on more covariates, as described in chapter 2.8.2.3, were performed. There are 155 data points (blocks of replicates) for which there is model 9 and 10 output. The system of making four categories for each covariate was used, as described in 2.8.2.2. The results of the sub-categorization are shown in Appendix L, Table L.2. For all subsets defined by that system which had at least 16 data points (10% threshold), correlation analysis was performed.

Model 9 had eleven subsets of data with lower P-value than for all the data. In all of these cases, the Spearman and Kendall tau correlations were higher for the subset than for all the data.

Table 3.17: Improved correlation between measured and predicted dermal absorption using model 9 for selected subsets of data.

Subset	log_10 P-value from Kendall test	Spearman correlation	Kendall tau	n
Conc>0.58	-8.2	0.53	0.37	116
HB≤7	-7.9	0.49	0.33	132
Charge>0.59, 4<HB≤7	-7.5	0.67	0.49	60
Charge>0.59, Conc>0.58, HB>4	-7.4	0.69	0.51	56
Conc>0.58, HB>4	-7.3	0.61	0.43	74
Conc>0.58, HB≤7	-7.2	0.52	0.36	106
HB≤7, logPow>1.25	-7.1	0.53	0.37	100
4<HB≤7	-7.1	0.57	0.41	80
Conc>0.58, logPow>1.25	-7.0	0.56	0.38	90
Conc>0.58, logPow≤4.11	-7.0	0.56	0.39	87
Conc>0.58, MW>249	-6.9	0.57	0.39	85
All data	-6.9	0.43	0.29	155

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In particular, sub-categorisation for concentration of the active substance and HB, the sum of hydrogen bond donors and acceptor numbers had the potential to improve correlation between measured and predicted values. Correlation was better when excluding the lowest category of test item concentration, i.e. very dilute mixtures, and when excluding active substances with low and high hydrogen bonding capacity. Figure 3.15 illustrates that the lowest and highest sub-ranges for HB have weaker association than the middle two ranges and removing the red points (lowest concentration range) clearly strengthens the associations, especially for the highest HB range. Restricting molecular weight to >249 g/mol may improve correlation (Spearman rho from 0.53 to 0.57) as this excludes preferentially small molecules which should typically have lower number of hydrogen bond donors / acceptors.

Model 10 had three subsets with lower P-value than for all the data and in every case the Spearman and Kendall tau correlations were higher than for all the data. As for Model 9, exclusion of data points associated with low test item concentrations (Table 3.18, Conc>0.58) and high number of hydrogen bond donors/acceptors (Table 3.18, HB≤7) improved correlation between measured and predicted data (Spearman rho from 0.41 to 0.50 and to 0.46, respectively). As illustrated in Figure 3.15, the model predicts much the same percentage dermal absorption for most cases with higher logPow. The previously described observation that high concentration yields lower measured absorption may be causative for improved correlation when excluding the data points relating to dilute test items.

It is considered that restriction of the model to certain ranges of logPow and/or concentration may optimise the predictive capacity of model 10. For example, the correlation in the subset with logPow≤4.11 and concentration above 0.58 g/L was 0.56 or 0.39 compared to 0.41 or 0.27 for the unrestricted dataset when performing Spearman or Kendall rank correlation analysis.

Table 3.18: Improved correlation between measured and predicted dermal absorption using Model 10 for selected subsets of data.

Subset	log_10 P-value from Kendall test	Spearman correlation	Kendall tau	n
Conc>0.58	-7.2	0.50	0.34	116
HB≤7	-7.0	0.46	0.31	132
Conc>0.58, logPoW≤4.11	-6.9	0.56	0.39	87
All data	-6.4	0.41	0.27	155

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

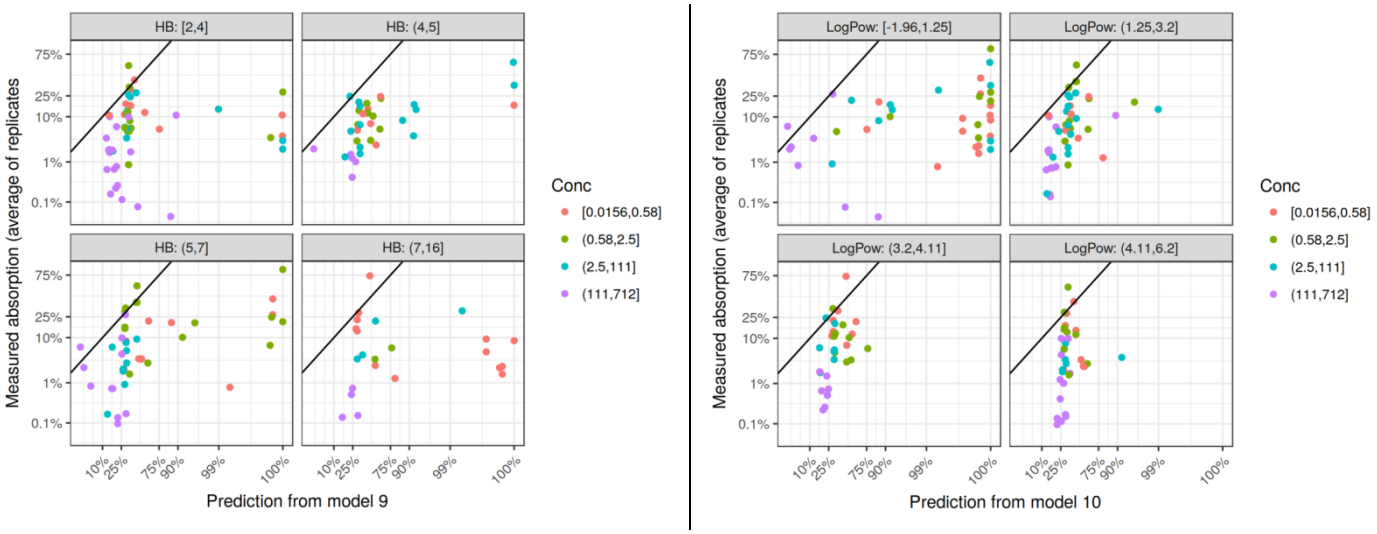


Figure 3.15: Relationship between the averaged experimental % Dermal Absorption (%DA) values and those predicted using: (left) model 9, subdivided into HB and Conc. (number of Hydrogen Bonds and concentration in mg/l, respectively) categories; (right) model 10, subdivided into LogPow and Conc. (concentration in mg/l) categories. Categories created as defined in Table L.2.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.6.1.3 Predictive capacity

In addition to the correlations analyses presented above (3.6.1.1 and 3.6.1.2), a more formal, empirical analysis of the predictive capacity of models 9 and 10 is shown here.

The empirical analysis of the predictive capacity for models 9 and 10 was conducted with averages of blocks of replicates. In comparison to model 7, the predictive capacity of models 9 and 10 is not as strong. This can be seen in Figure 3.16 where the difference between using the specific prediction and a randomly selected prediction is not as great as for model 7. This is due to the weaker correlation between predicted and measured absorption for these models. As previously noted (section 3.5.1.3), the empirical analysis performed in the present context faces some limitations. Even though it would be possible to complement the empirical approach with statistical modelling (as for model 7 in 3.5.2.3), the results outlined here do not suggest that this follow-up would not facilitate the understanding of the predictivity.

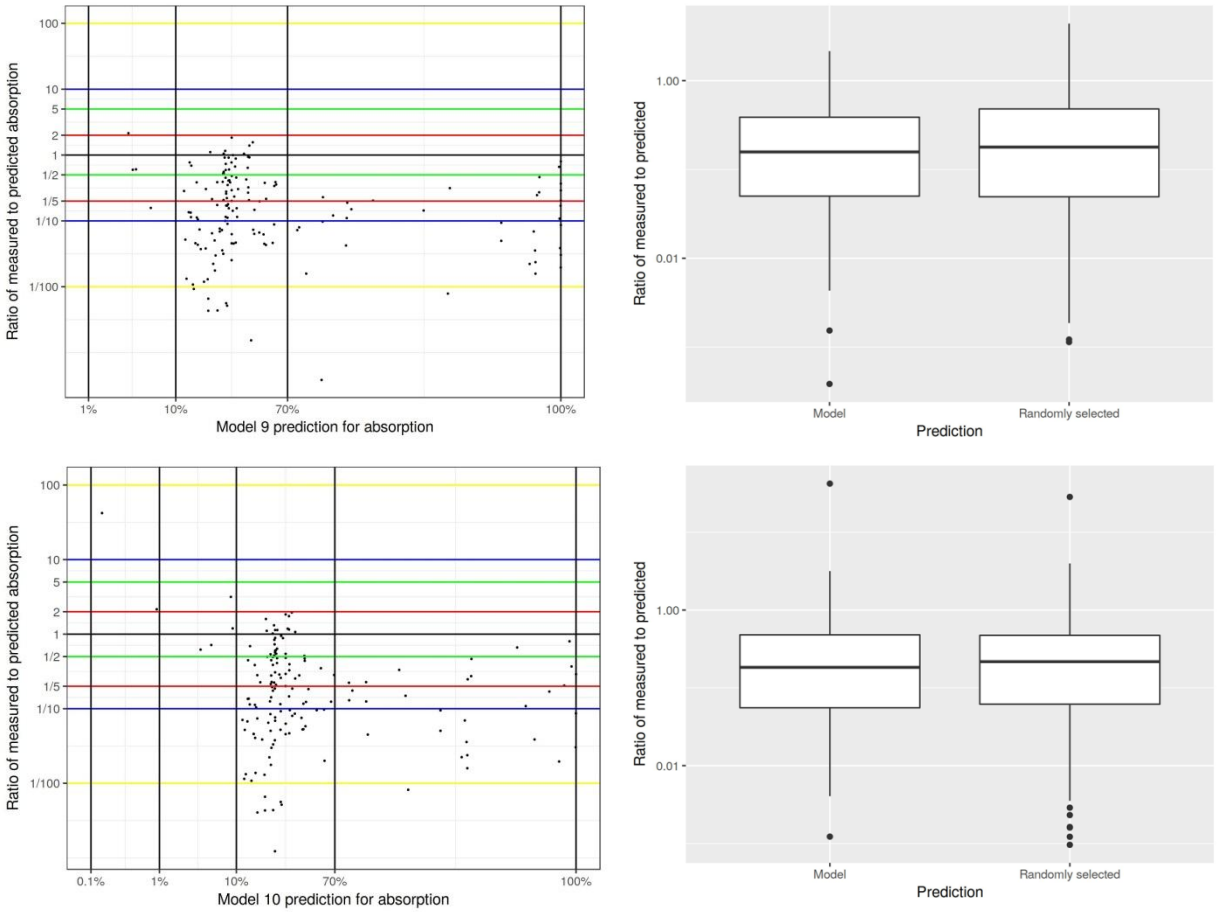


Figure 3.16: Left panels: Ratio of measured to predicted absorption versus predicted absorption for (top left) model 9 and (lower left) model 10. Vertical and horizontal lines highlight the categories used in Table 3.19 and Table 3.20. Right panels: Box and whisker plots of ratio of measured to predicted absorption using either (i) the specific prediction from model 9 (top right/ model 10 (lower right) for the measurement or (ii) a randomly selected prediction from model 9 (top right)/ model 10 (lower right).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.19: Table for model 9 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor.

	Model absorption prediction (depending on range of predicted value)		
Number of blocks:	4	150	1
Model prediction is:	**1-10%**	**10-100%**	**100%**
100 times lower than measured	0%	0%	0%
20 times lower than measured	0%	0%	0%
5 times lower than measured	0%	0%	0%
2 times lower than measured	25%	0%	0%
higher than measured	**75%**	**95%**	**100%**
2 times higher than measured	25%	**82%**	**100%**
5 times higher than measured	25%	**56%**	0%
10 times higher than measured	0%	39%	0%
100 times higher than measured	0%	6%	0%

Table 3.20: Table for model 10 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor.

	Model absorption prediction (depending on range of predicted value)		
Number of blocks:	2	123	30
Model prediction is:	**0.1-1%**	**1-10%**	**10-100%**
100 times lower than measured	0%	0%	0%
20 times lower than measured	**50%**	0%	0%
5 times lower than measured	**50%**	0%	0%
2 times lower than measured	**100%**	**25%**	0%
higher than measured	0%	**50%**	**92%**
2 times higher than measured	0%	0%	**80%**
5 times higher than measured	0%	0%	**54%**
10 times higher than measured	0%	0%	**39%**
100 times higher than measured	0%	0%	5%

Further considerations on uncertainties in prediction of models 9 and 10 are presented in chapter 3.8 focussing on regulatory relevance.

3.6.2 Model 11 – Atobe 2015

3.6.2.1 Correlations

Similar to most of the models examined here (see 3.5.1.1 for other models), no correlation between experimental and predicted %DA was apparent for model 11 (plots are to be found in Appendix K). This visual finding is also in line with the study of rank correlations (i) between measured individual replicates and corresponding %DA predictions and (ii) between measured logit transformed block aggregates, i.e. the mean of replicates within blocks and the corresponding averaged %DA prediction. The analysis resulted in low Spearman rho and Kendall tau rank correlation coefficients, ranging from 0.12 to 0.13 for Spearman rho and 0.08 to 0.09 for Kendall tau.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3.6.2.2 Influence of covariates

To explore potentially improved correlations, the relationships described above were broken down into subsets/plots[17] that additionally distinguish between:

i. dilution status (concentrate/in-use dilution)
ii. dilution status (concentrate/in-use dilution) and formulation types (organic solvent/ solid/water based/other)

The relationship plots (Appendix L, Figure L.1 & Figure L.2) and Spearman/Kendall tau rank correlations (Table 3.21a&b) revealed no apparent improvement in the correlation for any of the formulation types, justifying a more detailed statistical follow-up.

Table 3.21a: Spearman rank correlations between measured and predicted %DA for model 11, separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used.

Spearman Rank correlation	Dilutions	Concentrates	Solids	Water based	Organic solvent
Model 11	-0.01 (n=99)	0.4 (n=56)	0.08 (n=23)	0.06 (n=72)	0.13 (n=55)

Table 3.21b: Kendall tau rank correlations between measured and predicted %DA for model 11, separately for concentrates and dilutions and the three main formulation groups. Block aggregates were used.

Kendall tau Rank correlation	Dilutions	Concentrates	Solids	Water based	Organic solvent
Model 11	-0.01 (n=99)	0.28 (n=56)	0.06 (n=23)	0.03 (n=72)	0.08 (n=55)

3.7 Performance of Model Combinations

One of the objectives of the present study was to explore the possibility to improve predictions by model combination. Various combinations of the three strongest models 7, 9 and 10 were examined using the subset of the data for which the mixture models were applicable and predictions were available from both model 7 and the mixture models. As shown in Table 3.22, however, none of the combinations did a correlation that can be considered substantially better than the best individual model (here: model 9). It is considered unlikely that weighted averaging or use of a different form of mean would have a substantial impact on this outcome. However, this doesn't rule out the possibility that the models have different domains of strength, defined for example by certain physico-chemical properties of the active substance, or that particular combinations could perform more strongly in some restricted domain. Although some indications for better performance of individual models within such restricted domains were seen in the covariates analyses, the improvement of the quality of the correlation was moderate and at the price of significantly reducing the size of the corresponding data set. The number of blocks for the analysis in table 3.21 is already low and would become lower if subsets of the data were considered for models being combined and therefore this possibility was not pursued. In summary, this does currently not provide much encouragement for averaging.

[17] Sub-categorisation is in line with the proposals outlined in the revised EFSA Dermal Absorption guidance (2017)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.22: Various combinations of the three strongest models. Correlations were calculated for block aggregates using the subset of the data (n=44 blocks of replicates) for which model 7 was applied and for which the mixture models were applicable and provided output.

Combination of models	Correlation with measured absorption (Spearman rho)	Correlation with measured absorption (Kendall tau)
7 alone	0.48	0.32
9 alone	**0.57**	**0.41**
10 alone	0.54	0.38
7 and 9 averaged	0.59	0.41
7 and 10 averaged	0.58	0.39
9 and 10 averaged	0.57	0.41
7 and 9 and 10 averaged	0.59	0.43
minimum of 7 and 9	0.52	0.35
minimum of 7 and 10	0.52	0.34
minimum of 9 and 10	0.54	0.38
minimum of 7, 9 and 10	0.52	0.34
maximum of 7 and 9	0.59	0.41
maximum of 7 and 10	0.58	0.40
maximum of 9 and 10	0.57	0.41
maximum of 7, 9 and 10	0.59	0.41

3.8 Regulatory relevance of tested models

Member state consultation (see chapter 2.9) revealed several points of concern currently limiting the integration of *in silico* tools in the estimation of dermal absorption for pesticide risk assessment.

This included the relevance of the applicability domain of models for the prediction of dermal absorption of pesticides. Accordingly, both the literature assessment and the model selection took into consideration this issue. Only models with wide applicability domains - representing the majority of cases - were taken forward for detailed evaluation, while those limited to a certain class of chemicals were not considered potentially relevant to regulatory assessment of pesticides.

Another concern related to validation and transparency of model algorithms. Therefore, only publications that provided details on the training sets used for model development and the model's applicability domain were considered in detail in the literature review. Only well-defined and explicit algorithmic models and reconstructible machine learning models were evaluated further. Although proprietary *in silico* models were not excluded from the start, all of the eleven models that were eventually tested for predictivity are publicly available and easily accessible to the experts for use. The original published papers also provide statistical parameters to illustrate the goodness-of-fit, robustness and predictivity of the models as well as external validations. An additional asset is that for the tested models input data are accessible to a risk assessor or there are software tools available to calculate descriptors necessary to run the models. Thus, for all tested models, it should be feasible for an expert/risk assessor to both retrieve any of them and perform predictions.

Interestingly, a range of parameters that were regarded by experts as potentially influencing dermal absorption of pesticides were not taken into consideration by the models identified and evaluated here. This includes the substance/product classification with regard to skin irritation/corrosion or the effects of active substances and co-formulants on the *stratum corneum*, skin metabolism or surface tension of the active substance.

In this context, it is worth to note that the best performing models represented either mixture models (models 9 and 10), taking into account physicochemical properties of the co-formulants aggregated into a "mixture factor", or the complex single substance model no.7, which accounts for the type of

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

vehicle. Therefore, the outcome of the analysis may be related to this capacity. Information on the composition of the formulation was available only for a subset of the data, restricting the evaluation of the mixture models 9 and 10. Likewise, the manual effort associated with implementation of model 7 restricted the dataset for evaluation of this approach as well, although representativeness was assured.

Model 7 or "NIOSH model" is a complex spreadsheet-based model for estimating bioavailability of a chemical from dermal exposure. The model takes into consideration differences in skin parameters between species (e.g. rat or human), anatomical site, temperature (of skin) and hydration of *stratum corneum*. It also takes into consideration the experimental conditions e.g. state of occlusion (occlusive, semi-occlusive or non-occlusive) or preparation of the skin sample (epidermal sheet, dermatomed skin or full-thickness skin) and duration of exposure. Further, it also considers some physicochemical properties of the formulations (e.g. pH) and their in-use dilutions, the active substance (e.g. physical state, molecular weight, lipid/water partition coefficient, ionization, water solubility, pKa), the vehicle(s) and other co-formulants (e.g. solubility, volatility). Also, the partition partition coefficient of the compound of interest between vehicle and water needs to be known or estimated. For the purpose of the project, the vehicle was either treated as water or octanol depending on formulation type and dilution status allowing the use of Ko/w. Since a common template is available for this model, competent authorities and applicants could, in principle, perform the predictions in a harmonised manner. The wide variety of exposure and computation settings that are offered by the model, requires users to make assumptions before performing any calculation, as also seen in the present study and extensively described in 2.7.1.7. As these may influence model prediction, agreement on these choices would be necessary and a practical guide should be available to avoid other pitfalls.

Models 9 and 10 are both algorithm-based, mixture models that predict k_p by considering the octanol/water partition coefficient and the molecular weight of the active substance as well as the physical-chemical properties of the mixture based on its components (mixture factor). In model 9 the mixture factor is related to the topological polar surface area of the mixture, whereas in model 10 it is related to the number of hydrogen bond acceptor of the mixture. For development of models 9 & 10, test compounds were diluted in different vehicles by creating in total 24 different vehicle mixtures from ethanol, methylnicotinic acid, propylenglycol, sodium laureth sulfate and water. Therefore, the training dataset can be considered to include aqueous and organic based formulations. There is no information in the publication regarding the use of solids. No other physicochemical properties of the formulation are considered. There is also no consideration of the complexity of skin or the experimental conditions in either of models 9 or 10. Accordingly, there may be various options for refinement of these models.

The overall outcome of the performance of models 7, 9 and 10, as presented in detail in chapters 3.5 and 3.6, has been the basis for further discussion on their potential integration within the regulatory framework. With respect to the outcome of the survey with the MSCA experts (see chapter 3.1), the following points have been further elaborated:

The possibility to improve predictivity by restriction to subsets of data, i.e. by narrowing the applicability domain, was investigated. This included proposals to consider subsets for (a) concentrates and dilutions or (b) different formulation types (three main formulations groups have been considered, i.e. solids, water-based and organic solvent). In some cases, predictivity was slightly but not strikingly improved. Similarly, the consideration of testing the skin type (full thickness, dermis, epidermis, scissors) as a covariate was not successfull. For model 7, no new subsets could be produced with regard to SkinType. The correlation outcome for models 9 and 10, when subsetting the data with respect to SkinType, was not stronger compared to the non-subsetted dataset (data not included).

Since model 7 performed slightly better than models 9 and 10, more extensive analyses were performed. When pKa was used for subgrouping of substances in the statistical analysis of this model, it was revealed that there might be a difference in predictivity for substances with pKa ≤ 9.4 and

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

those with pKa > 9.4. It is noted that during implementation of model 7, pKa data (either measured or calculated) were not included in the prediction, since pKa is one of the optional input parameters in the model (Dancik *et al.,* 2013). It would thus be interesting to further investigate the influence of pKa and other optional physical-chemical properties, such as water solubility, density, vehicle solubility, non-ionised unbound fractions, on the predictivity of model 7.

Finally, a Bayesian random effects model (BREM) was used to adjust the prediction of model 7 in order to achieve better predictive performance as described in chapter 3.5.2.3. Using this approach, adjustment for the tendency for predictions of low absorption to under-predict and for predictions of high absorption to over-predict was made.

Within the established tiered approach for assessment of dermal absorption of pesticides, modelling would allow a refinement over default values only when reliable values below the applicable defaults are predicted. Accordingly, predictions from models 7, 9 and 10, taking into account the degree of over- und under-predictions, were compared to default values as presented in the new EFSA Guidance (EFSA, 2017). First, an analysis was conducted to calculate the percentage of experimental blocks for which each of models 7, adjusted model 7 as well as models 9 and 10 gives predictions below the respective default value as set in the Table 2 in the EFSA GD (2017). Results are summarised in Tables 3.23 and 3.24 below. The possibility to also use assessment factors (AF) of 5 and 10 to the adjusted predictions from model 7 was further investigated. The assessment factor was applied by first calculating the adjusted prediction from model 7 and then multiplying the result by the assessment factor. If the result exceeds 100%, it is is capped at 100%. Continuing the earlier example, if model 7 predicts 0.5% absorption, the adjusted prediction is 2.1% which becomes $5 \times 2.1\% = 10.5\%$ when an assessment factor of 5 is applied or 21% if the assessment factor is 10.

It was shown that in case of the adjusted model 7 (Table 3.23) 100% of the predicted values are less than the respective EFSA (2017) default values for both the concentrate and the dilution for all formulation types. Thus, if acceptable, this model would allow refinement of default values in essentially all cases. For the other models, the percentage of cases were refinement may be possible was dependent on the formulation group, the dilution status and the application of an assessment factor.

Table 3.23: For model 7 and adjusted model 7 with and without assessment factor (AF): percentage of blocks for which prediction is less than the EFSA (2017) default value based on formulation group and concentrate/dilution status.

Formulation group	Concentrate or Dilution	Default value	Model				No of blocks
			7	7 adj.	7 adj., AF=5	7 adj., AF=10	
Organic solvent	Dilution	70%	58%	100%	54%	27%	52
Organic solvent	Concentrate	25%	95%	100%	91%	82%	22
Other	Dilution	70%	50%	100%	50%	25%	8
Other	Concentrate	25%	100%	100%	100%	25%	4
Solid	Dilution	50%	71%	100%	46%	20%	35
Solid	Concentrate	10%	100%	100%	58%	38%	26
Water based	Dilution	50%	85%	100%	71%	38%	48
Water based	Concentrate	10%	97%	100%	67%	58%	36

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.24: For models 9 and 10: percentage of blocks for which prediction is less than the EFSA (2017) default value based on formulation group and concentrate/dilution status.

Formulation group	Concentrate or Dilution	Default value	Model		No of blocks
			9	10	
Organic solvent	Dilution	70%	80%	80%	30
Organic solvent	Concentrate	25%	40%	44%	25
Other	Dilution	70%	100%	100%	2
Other	Concentrate	25%	100%	100%	1
Solid	Dilution	50%	30%	43%	23
Solid	Concentrate	10%	N/A	N/A	0
Water based	Dilution	50%	56%	65%	43
Water based	Concentrate	10%	7%	7%	29

Furthermore, from a regulatory perspective, quantification the uncertainty in the predicted dermal absorption values for any recommended model is considered to be crucial, especially taking into account that the predictivity found in most cases was regarded not better than moderate.

The following tables present how frequently model 7, 9 and 10 as well as adjusted model 7 exceed measurement by specified factor depending on the range of predicted values. These tables complement those shown in chapters 3.5 and 3.6 by specifying frequencies for the ranges below the 70% default value of the EFSA (2017) Guidance. Only those cases are included, where use of the predicted value would result in a refinement of the default value to a lower estimate.

Table 3.25: Table for model 7 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor. Data points (blocks) are restricted to those for which predictions are lower than the applicable EFSA (2017) default values.

	Model absorption prediction (depending on range of predicted value)			
Model prediction is:	**0.01- <0.1%**	**0.1- <1%**	**1- <10%**	**10- <70%**
Number of blocks below applicable default:	28	30	76	37
100 times lower than measured	7%	10%	0%	0%
20 times lower than measured	**75%**	23%	1%	0%
5 times lower than measured	**86%**	**53%**	9%	3%
2 times lower than measured	**100%**	**73%**	28%	11%
higher than measured	0%	10%	**54%**	**81%**
2 times higher than measured	0%	3%	**39%**	**49%**
5 times higher than measured	0%	3%	13%	24%
10 times higher than measured	0%	3%	7%	14%
100 times higher than measured	0%	3%	1%	0%

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.26: Table for model 9 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor. Data points (blocks) are restricted to those for which predictions are lower than the applicable EFSA (2017) default values.

	Model absorption prediction (depending on range of predicted value)	
Model prediction is:	**1- <10%**	**10- <70%**
Number of blocks below applicable default:	**4**	66
100 times lower than measured	0%	0%
20 times lower than measured	0%	0%
5 times lower than measured	0%	0%
2 times lower than measured	25%	0%
higher than measured	**75%**	**91%**
2 times higher than measured	25%	**67%**
5 times higher than measured	25%	**35%**
10 times higher than measured	0%	18%
100 times higher than measured	0%	0%

Table 3.27: Table for model 10 of how frequently model exceeds measurement by specified factor or measurement exceeds model by specified factor. Data points (blocks) are restricted to those for which predictions are lower than the applicable EFSA (2017) default values.

	Model absorption prediction (depending on range of predicted value)		
Model prediction is:	**0.1- <1%**	**1- <10%**	**10- <70%**
Number of blocks below applicable default:	2	4	72
100 times lower than measured	0%	0%	0%
20 times lower than measured	**50%**	0%	0%
5 times lower than measured	**50%**	0%	0%
2 times lower than measured	**100%**	25%	0%
higher than measured	0%	**50%**	**85%**
2 times higher than measured	0%	0%	**64%**
5 times higher than measured	0%	0%	**33%**
10 times higher than measured	0%	0%	22%
100 times higher than measured	0%	0%	0%

When considering the respective results of the adjusted model 7 presented in Table 3.28 below, the model prediction is at least 2 times lower than the measured one for all cases examined when the range of the predicted values is low, i.e. 0.01-0.1%. When the predicted values are in the range of 10 to 70% then for 52% of the cases the predicted value is higher than the measured one.

In consequence, for regulatory purposes, the use of the predicted values from models 7, 9, 10 and adjusted model 7 in general may not be acceptable due to a notable risk of under-prediction. However, if circumstances (conditions) can be identified, under which the predicted value is higher than the measured one but lower than the respective recommended default value (EFSA, 2017), then the possibility to use the predicted value for a refinement could be further considered.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.28: How frequently adjusted prediction from model 7 exceeds measurement by specified factor or measurement exceeds model by specified factor. Data points (blocks) are restricted to those for which predictions are lower than the applicable EFSA (2017) default values.

	Adjusted Model absorption prediction (depending on range of predicted value)			
Model prediction is:	**0.01- <0.1%**	**0.1- <1%**	**1- <10%**	**10- <70%**
Number of blocks below applicable default:	7	34	122	67
100 times lower than measured	0%	0%	0%	0%
20 times lower than measured	**43%**	9%	2%	0%
5 times lower than measured	**86%**	12%	7%	0%
2 times lower than measured	**100%**	29%	22%	16%
higher than measured	0%	**56%**	**56%**	**52%**
2 times higher than measured	0%	**35%**	**39%**	22%
5 times higher than measured	0%	3%	18%	6%
10 times higher than measured	0%	0%	7%	1%
100 times higher than measured	0%	0%	2%	0%

Thus, an attempt was made to conduct further analysis of model 7 outcomes in order to address this issue and provide risk assessors and risk managers as many information as possible in order to further consider whether the uncertainty quantified could be tolerated for regulatory purposes. Respective analyses could also be performed for models 9 and 10. For model 7, it was observed that when the predicted values are multiplied with an assessment factor of 5 or 10, the percentage of cases where the resulting predicted value is higher that the measured one increases to 94% and 96%, respectively for the range of 10 to 70% modelled dermal absorption (Table 3.29 & Table 3.30). Comparing these results with the presented uncertainty in prediction in the tables above, it is notable that by applying an assessment factor of at least 5 the chance of under-prediction is significantly reduced while still allowing refinement over the default values of EFSA (2017).

Table 3.29: How frequently adjusted prediction from model 7, with assessment factor (AF) of 5 applied, exceeds measurement by specified factor or measurement exceeds model by specified factor. Data points (blocks) are restricted to those for which predictions are lower than the applicable EFSA (2017) default values.

	Model absorption prediction (depending on range of predicted value)			
Model prediction is:	**0.01- <0.1%**	**0.1- <1%**	**1- <10%**	**10- <70%**
Number of blocks below applicable default:		9	55	80
100 times lower than measured		0%	0%	0%
20 times lower than measured		22%	0%	0%
5 times lower than measured		**33%**	11%	0%
2 times lower than measured		**33%**	11%	0%
higher than measured		**33%**	**87%**	**92%**
2 times higher than measured		11%	**73%**	**78%**
5 times higher than measured		0%	**56%**	**45%**
10 times higher than measured		0%	33%	30%
100 times higher than measured		0%	2%	1%

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 3.30: How frequently adjusted prediction from model 7, with assessment factor of 10 applied, exceeds measurement by specified factor or measurement exceeds model by specified factor**.** Data points (blocks) are restricted to those for which predictions are lower than the applicable EFSA (2017) default values.

	Model absorption prediction (depending on range of predicted value)			
Model prediction is:	**0.01- <0.1%**	**0.1- <1%**	**1- <10%**	**10- <70%**
Number of blocks below applicable default:		7	34	49
100 times lower than measured		0%	0%	0%
20 times lower than measured		0%	0%	0%
5 times lower than measured		29%	0%	0%
2 times lower than measured		**43%**	9%	6%
higher than measured		**57%**	**91%**	**94%**
2 times higher than measured		14%	**88%**	**90%**
5 times higher than measured		0%	**71%**	**65%**
10 times higher than measured		0%	**56%**	41%
100 times higher than measured		0%	0%	0%

The results summarised in Tables 3.28 to 3.30 are illustrated in the following figures:

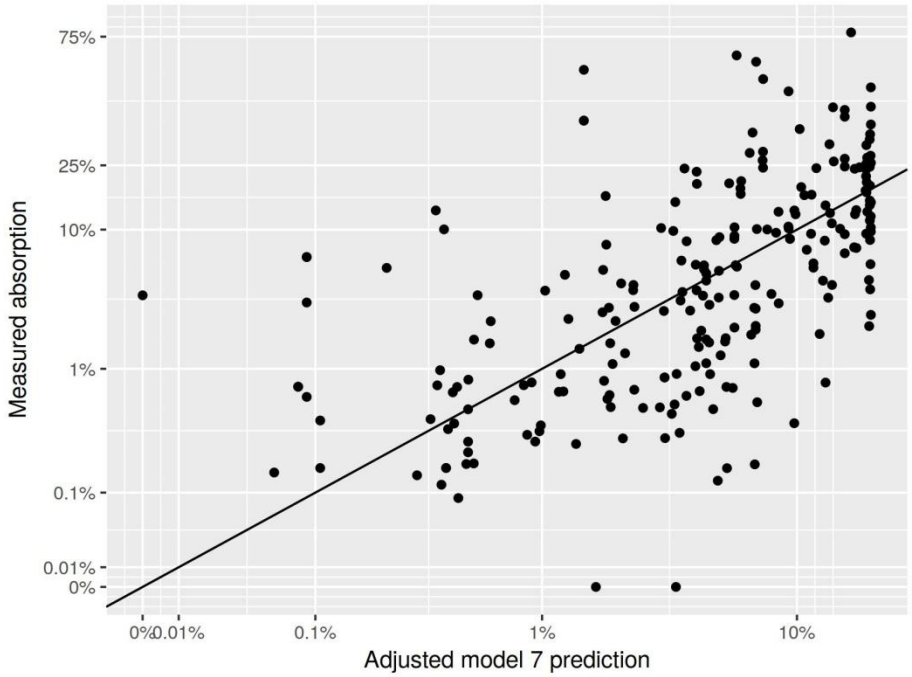


Figure 3.17: Plot of measured absorption versus adjusted prediction from model 7. Black line shows where they are equal.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

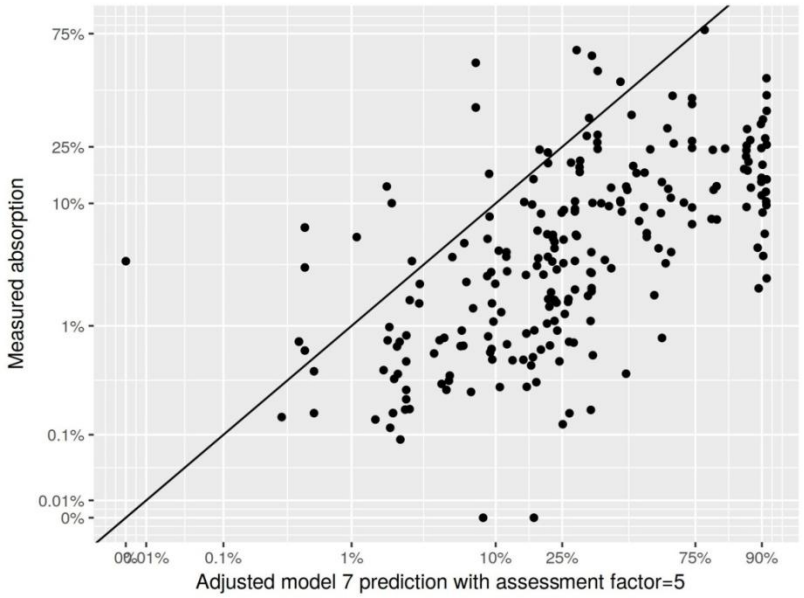


Figure 3.18: Plot of measured absorption versus adjusted prediction from model 7 having multiplied the adjusted prediction by an assessment factor of 5. Values above 100% are capped at 100%. Black line shows where they are equal.

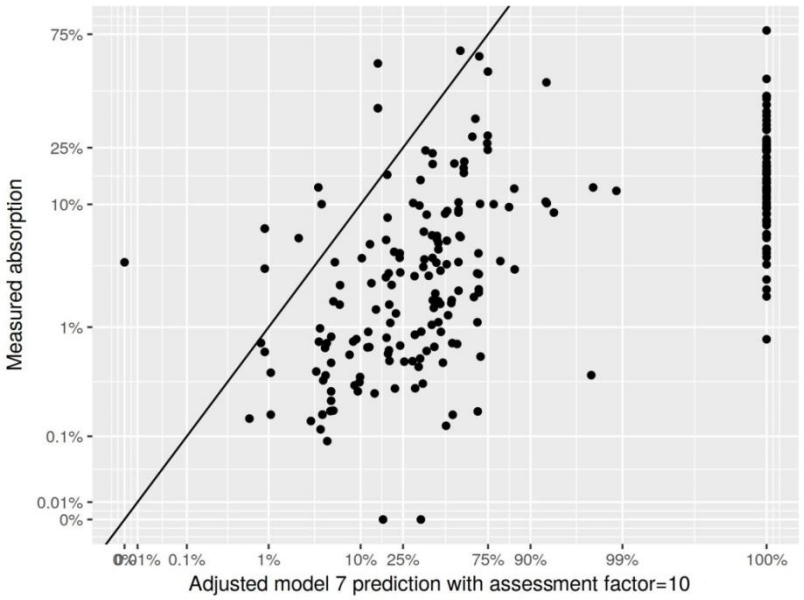


Figure 3.19: Plot of measured absorption versus adjusted prediction from model 7 having multiplied the adjusted prediction by an assessment factor of 10. Values above 100% capped at 100%. Black line shows where they are equal.

Overall, considering the result from model implementation and statistical analysis from a regulatory perspective, it is recommended to further investigate the possibility of using the adjusted model 7 (NIOSH model) as a Tier 2 tool to estimate the dermal absorption value with a defined level of uncertainty in the absence of adequate experimental data when the predicted values are in the range of 10 to 70%,. The outcome of the above presented analyses shows that, if using the adjusted model

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

7 as been used during the implementation stage of this project, the application of an assessment factor of at least 5 is required to increase acceptability for regulatory purposes. In addition, it should be noted that the use of the adjusted model 7 may be acceptable in the range of 10 – 70% dermal absorption only, i. e. within the range of the default values of the EFSA Guidance on dermal absorption (EFSA, 2017). The uncertainty of prediction, or better, the risk of under-prediction might be too high for lower predicted dermal absorption values (0.01 – 10%).

The same considerations that were exemplarily presented for model 7 could be applied to models 9 and 10. However, as noted above, implementation of models 9 and 10 requires access to confidential data on the composition of the formulation, which in many cases is limited with regard to the exact molecular composition. This is not required for model 7. In addition, the strength of correlation between predicted and measured data as a prerequisite for practical implementation was regarded less convincing for models 9 & 10 compared to model 7.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

4 Conclusions

Consultation with Member State Competent Authorities (MSCAs) on the practical applicability of *in silico* tools was used as the starting point to define regulatory needs and identify technical questions that should be addressed in this project. The result of the survey showed that most MSCA experts may lack the technical knowledge for using available dermal absorption prediction tools reliably. Provision of clear recommendations by regulatory bodies (e.g. EFSA) would facilitate wider acceptance and use of *in silico* tools for that purpose. Two major components were identified as essential for such recommendations: (i) clear definition of the applicability of the model to be used, considering the complexity of agrochemical formulations and realistic exposure scenarios, and (ii) the possibility to transparently address uncertainties associated with the %DA estimations.

A systematic literature review was performed with the aim to identify relevant *in silico* tools whichrevealed an unexpectedly high number of 142 original research papers plus 43 review papers describing a total of 288 models for prediction of dermal absorption with potential relevance for pesticide risk assessment. Following scientific review of these publications, the significant number of 61 algorithm-based and 14 machine-learning models for prediction of dermal absorption of single substances and their dilutions in water were identified, while there were 29 algorithm-based and 7 machine-learning models aimed at prediction of dermal absorption of single substances from vehicles representing mixtures including solvents other than water.

When selecting models for implementation and detailed statistical evaluation in a stepwise procedure, the availability of input data or software tools to calculate descriptors necessary to run the model was identified as a major bottleneck. The availability of measured or predicted input data can severely limit the range of models that can practically be used. Also, it was noted during model implementation that strategic choices have to be made, e.g. to the inclusion of lag-time or the use of calculated versus experimental water solubility, showing that the same model may be used (or interpreted) differently between groups. Effectively, this can be generating subversions of one model or model combination in practice. Also, when comparing measured vs. predicted physicochemical parameters, it was noted that these can deviate substantially. This may potentially have a significant impact on the outcome of model predictions for dermal absorption. Mining of data for mixture composition, including pesticidal products and in-use dilutions thereof, was very labor intensive and, based on the experience made in this project, will usually not result in full description of the mixture with information on content, structure, physicochemical or quantum-chemical descriptors for some of the components often remaining unknown. Thus, mixture descriptors usually represent best estimates only. While algorithmic models are typically easily implemented in spreadsheet software, models on artificial neural networks require more sophisticated solutions that may not be available at evaluating authorities. Even if reconstructed, artificial neural networks may differ with respect to prediction outcome from the original ANN.

Most models predicted either permeation coefficients or maximum flux values and thus need to be combined with models extrapolating to percentage dermal absorption – the output parameter currently used in pesticide risk assessment. Within the project, this was done by feeding the predicted k_p values into the DAME model (Buist 2010) except for model 7. Although the authors considered DAME as the best available tool for the purpose, there are other legitimate approaches which could not be explored here, however. In this context, it was noted that predicted %DA was the sum of % predicted in the receptor fluid and the *stratum corneum* with no consideration of the barrier function of remaining skin layers, except for model 7, where the model allowed to predict %DA as the sum of % absorbed systemically, in the *stratum corneum*, viable epidermis, and dermis.

Plots depicting the relationship between predicted and measured %DA, did not reveal any correlations between the two variables for most models. Weak to moderate associations between predicted and measured absorption, as demonstrated by the respective graphical representations and confirmed by Spearman and Kendall tau rank correlations analysis, respectively, were detected for model 7 (single substance model) and for models 9 & 10 (mixture models). This finding suggests that models 7, 9 and 10 were worthy of further investigation and that the other models are less likely to be useful

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

predictively. The fact that moderate predictivity was achieved with models 9 and 10 suggests that lack of correlation for the poorly predicting models cannot be primarily attributed to the use of DAME. However, the best performing model (model 7), published by NIOSH and most recently described by Dancik *et al.* (2013), is one of the few models predicting percentage absorption directly, without the use of another model. With regard to models 9 and 10, it is worth to note that these were conceived by Riviere & Brooks (2007) based on the well-known model by Potts&Guy (1992) and do - in this analysis - indeed provide an improvement over the "gold standard".

In addition, it may be concluded that the improved performance of models 7, 9 and 10 may be related by the capability of model 7 to take into account (semi-quantitatively) some characteristic of the vehicle mixtures, while models 9 & 10 were designed to address vehicles representing mixtures. Selection bias resulting from the restriction of the dataset to those formulations, for which the composition could be identified, was ruled out. However, it remains obscure why the other mixture model published by Atobe *et al.* (2015) did not perform well. Reconstruction of this ANN was shown to be successful, when using the original dataset.

In a search for approaches to further improve predictivity of models 7, 9 & 10, subsets of the data were generated defined by different ranges of selected covariate values for which there might be stronger correlations. Although there was slightly better correlation for some subsets defined by certain pKa ranges (model 7) or the formulation type (e.g. water based), this exercise did not reveal any subsets with very strong correlations. This does not mean that they necessarily do not exist but does mean that they are likely to be difficult to detect due to the need to avoid spurious false positives when checking a large number of possible subsets.

A Bayesian Random Effects Model (BREM) allowed to analyse how over- and under-prediction depends on the predicted level of absorption for model 7. It was also successful in showing how the output of model 7 might be adjusted to reduce the magnitude of over- and under-prediction in those ranges. However, it was concluded that no statistical methodology, including the BREM, would be capable of overcoming completely the problems caused by relative weak correlation between predictions and measurements.

Overall, considering the result from model implementation and statistical analysis, it is recommended to further investigate the possibility of using the adjusted model 7 (NIOSH model) as a Tier 2 tool to estimate the dermal absorption value in the absence of adequate experimental data when the predicted values are in the range of 10 to 70%, with a defined level of uncertainty. However, the outcome of the above presented analyses shows that, if using the adjusted model 7, the application of an assessment factor of at least 5 is required to increase acceptability for regulatory purposes. In addition, it should be noted that the use of the adjusted model 7 is only acceptable in the range of 10 – 70% dermal absorption, i.e. the default values of the EFSA Guidance on dermal absorption (2017). The uncertainty of prediction, or better the risk of under-prediction might be too high in lower predicted dermal absorption values (0.01 – 10%).

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

5 Recommendations

- Clearly, further improvement of models for *in silico* prediction of dermal absorption is required to achieve wide-spread acceptance. It is recommended to further investigate the influence of physical properties defined as optional by Dancik et al (2013) such as pKa, water solubility, density, vehicle solubility, non-ionised unbound fractions, on the predictivity of model 7. Since the NIOSH model may use k_p as an input parameter, it could be investigated whether combining the NIOSH model with the selected models predicting k_p or J_{max} would yield a better performing predictive tool. Notably, none of the eleven models tested takes into consideration the substance/product classification with regard to skin irritation/corrosion, co-formulant's effects on the *stratum corneum*, metabolism or surface tension of the pesticide; elements that the MSCA experts would expect to have an impact on dermal absorption according to the survey.

- To facilitate model development, existing dermal absorption datasets, both public and private, should be combined, curated and filtered to build a large homogeneous and consistent dataset. The on-line available EDETOX-database[18] provides a good structure to gather these data and already a fair number of curated k_p data. In order to increase the probability that the newly developed model will include good predictions of pesticides, also measured k_p values of pesticides should be collected. However, confidentiality issues may impact negatively on the representation of data for commercially relevant pesticide formulations in such a database.

- Reporting identity of co-formulants in pesticidal products should be re-considered. When collecting information on mixture composition for the purpose of this project, it was noted that from the information provided in the confidential parts of the Registration Reports, the factual chemical composition could only be approximated. Frequently, the chemical structure of a co-formulant can either not be unambiguously identified or a co-formulant represents a mixture of substances itself with often unclear structure or content in the mixture.

- Considering the result from statistical analysis, it is recommended to further investigate the possibility of using the adjusted model 7 (NIOSH model) as a Tier 2 tool for estimation of dermal absorption in the absence of adequate experimental data. However, the outcome of the above presented analyses shows that, if using the adjusted model 7, the application of an assessment factor of at least 5 is required to increase acceptability for regulatory purposes. In addition, it should be noted that the use of the adjusted model 7 is only acceptable in the range of 10 – 70% dermal absorption, i. e. the default values of the EFSA Guidance on dermal absorption (2017). The uncertainty of prediction, or better the risk of under-prediction might be too high in lower predicted dermal absorption values (0.01 – 10%).

- Adjustment of data requirements may need to be considered when (Q)SAR models are to be used at larger scale. Depending on the individual (Q)SAR model, certain structural, physicochemical, or quantum-chemical descriptors are needed as input parameters to run the model or refine the prediction. However, in some cases, these parameters may not be covered by the legal data requirements and may not be available. This applies in particular to co-formulants. In this context, it should be considered to validate and make available tool for prediction of substance properties ((Q)SPR tools), needed to calculate input parameters for prediction models.

[18] http://edetox.ncl.ac.uk

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

- If recommending (Q)SAR for pesticide risk assessment including but not limited to prediction of dermal absorption, it should be considered to provide detailed technical guidance. Various potential pitfalls were noted during this project, starting already at the stage of collection, calculation or prediction of input parameters. Depending on the approach to obtain an input parameter, very different values can be obtained, potentially impacting the output. In addition, handling or interpretation errors may easily occur to the unfamiliar risk assessor while using (Q)SAR / (Q)SPR tools. As described above, choices to be made during implementation may effectively lead to establishing different subversions of one model.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

References

Abraham, M. H. and F. Martins (2004). "Human skin permeation and partition: general linear free-energy relationship analyses." J Pharm Sci 93(6): 1508-1523.

Abraham, M. H., Chadha, H. S., & Mitchell, R. C. (1995). The Factors that Influence Skin Penetration of Solutes. Journal of Pharmacy and Pharmacology, 47(1), 8-16. doi:10.1111/j.2042-7158.1995.tb05725.x

Abraham, M. H., F. Martins and R. C. Mitchell (1997). "Algorithms for skin permeability using hydrogen bond descriptors: The problem of steroids." Journal of Pharmacy and Pharmacology 49(9): 858-865.

Abraham, M. H., Ibrahim, A., Zissimos, A. M., Zhao, Y. H., Comer, J., & Reynolds, D. P. (2002). Application of hydrogen bonding calculations in property based drug design. Drug Discov Today, 7(20), 1056-1063.

Anissimov, Y. G., O. G. Jepps, Y. Dancik and M. S. Roberts (2013). "Mathematical and pharmacokinetic modelling of epidermal and dermal transport processes." Advanced Drug Delivery Reviews 65(2): 169-190.

Atobe, T., Mori, M., Yamashita, F., Hashida, M., & Kouzuki, H. (2015). Artificial neural network analysis for predicting human percutaneous absorption taking account of vehicle properties. Journal of Toxicological Sciences, 40(2), 277-294.

Auguie, B. (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Bradley, J. C., Abraham, M. H., Acree, W. E., Jr., & Lang, A. S. (2015). Predicting Abraham model solvent coefficients. Chem Cent J, 9, 12. doi:10.1186/s13065-015-0085-4

Buchwald, P., & Bodor, N. (2001). A simple, predictive, structure-based skin permeability model. J Pharm Pharmacol, 53(8), 1087-1098.

Buist, H. E., 2016. Dermal absorption and toxicological risk assessment: pitfalls and promises. PhD thesis, Wageningen University. Wageningen University, Wageningen, the Netherlands.

Buist, H. E., van Burgsteden, J. A., Freidig, A. P., Maas, W. J. M., & van de Sandt, J. J. M. (2010). New in vitro dermal absorption database and the prediction of dermal absorption under finite conditions for risk assessment purposes. Regulatory Toxicology and Pharmacology, 57(2-3), 200-209. doi:10.1016/j.yrtph.2010.02.008

Dancik, Y., Miller, M. A., Jaworska, J., & Kasting, G. B. (2013). Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. Advanced Drug Delivery Reviews, 65(2), 221-236. doi:10.1016/j.addr.2012.01.006

EFSA, 2012. Guidance on dermal absorption. EFSA Journal 2012;10(4):2665.

EFSA, 2017. Guidance on dermal absorption. EFSA Journal 2012;15(6):4873.

Ertl, P., Rohde, B., & Selzer, P. (2000). Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem, 43(20), 3714-3717.

Flynn, G. L.. Physicochemical determinants of skin absorption. In Principles of Route-to-Route Extrapolation for Risk Assessment, T. R. Gerrity and C. J. Henry (eds.), Elsevier, New York, 1990, pp. 93-127.

Frasch, H. F. (2002). A random walk model of skin permeation. Risk Anal, 22(2), 265-276.

Fu, X. C., Ma, X. W., & Liang, W. Q. (2002). Prediction of skin permeability using an artificial neural network. Pharmazie, 57(9), 655-656.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Geinoz, S., Guy, R.H., Testa, B., Carrupt, P.-A., 2004. Quantitative Structure-Permeation Relationships (QSPeRs) to Predict Skin Permeation: A Critical Evaluation. Pharm. Res. 21, 83–92

Ghafourian, T., E. G. Samaras, J. D. Brooks and J. E. Riviere (2010a). "Modelling the effect of mixture components on permeation through skin." International Journal of Pharmaceutics 398(1-2): 28-32.

Ghafourian, T., E. G. Samaras, J. D. Brooks and J. E. Riviere (2010b). "Validated models for predicting skin penetration from different vehicles." European Journal of Pharmaceutical Sciences 41(5): 612-616.

Gute, B. D., G. D. Grunwald and S. C. Basak (1999). "Prediction of the dermal penetration of polycyclic aromatic hydrocarbons (PAHs): a hierarchical QSAR approach." SAR and QSAR in environmental research 10(1): 1-15.

Guy, R. H., & Potts, R. O. (1992). Structure- permeability relationships in percutaneous penetration. Journal of Pharmaceutical Sciences, 81(6), 603-604. doi:10.1002/jps.2600810629

Guy, R. H., & Potts, R. O. (1995). A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. Pharmaceutical Research, 12(11), 1628-1633. doi: 10.1023/A:1016236932339

Højsgaard, S. & Halekoh U. (2016). doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities. R package version 4.5-15. https://CRAN.R-project.org/package=doBy

Hostýnek, J. J. and Magee, P. S. (1997), Modelling In Vivo Human Skin Absorption. Quant. Struct.- Act. Relat., 16: 473-479. doi:10.1002/qsar.19970160606

Hou, T. J., Xia, K., Zhang, W., & Xu, X. J. (2004). ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. Journal of Chemical Information and Computer Sciences, 44(1), 266-275. doi:10.1021/ci034184n

Hui X, Wester RC, Magee PS, Maibach HI (1995) Partitioning of chemicals from water into powdered human stratum corneum (Callus): A model study. In Vitro Toxicology: Journal of Molecular and Cellular Toxicology 8(2):159-167

Jain, N. and S. H. Yalkowsky (2001). "Estimation of the aqueous solubility I: application to organic nonelectrolytes." J Pharm Sci 90(2): 234-252.

Komsta, L. and Novomestky, F. (2015). moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14. https://CRAN.R-project.org/package=moments

Kronmal, R. (1993). Spurious correlation and the fallacy of the ratio standard revisited. Journal of the Royal Statistical Society. Series A (Statistics in Society), 156(3), 379-392.

Magnusson, B. M., Anissimov, Y. G., Cross, S. E., & Roberts, M. S. (2004). Molecular size as the main determinant of solute maximum flux across the skin. J Invest Dermatol, 122(4), 993-999. doi:10.1111/j.0022-202X.2004.22413.x

McKone, T. E., Howd, R. A., 1992. Estimating dermal uptake of nonionic organic chemicals from water and soil: I. Unified fugacity-based models for risk assessments. Risk Anal. 12, 543-57.

Milewski, M., & Stinchcomb, A. L. (2012). Estimation of maximum transdermal flux of nonionized xenobiotics from basic physicochemical determinants. Molecular Pharmaceutics, 9(7), 2111-2120. doi:10.1021/mp300146m

Mitragotri, S., Y. G. Anissimov, A. L. Bunge, H. F. Frasch, R. H. Guy, J. Hadgraft, G. B. Kasting, M. E. Lane and M. S. Roberts (2011). "Mathematical models of skin permeability: an overview." Int J Pharm 418(1): 115-129.

Moss, G.P., Cronin, M.T., 2002. Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data. Int. J. Pharm. 238, 105–109.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Moss, G.P., Sun, Y., Wilkinson, S.C., Davey, N., Adams, R., Martin, G.P., Prapopopolou, M., Brown, M.B., 2011. The application and limitations of mathematical modelling in the prediction of permeability across mammalian skin and polydimethylsiloxane membranes. J. Pharm. Pharmacol. 63, 1411–1427.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. J Cheminform, 3, 33. doi:10.1186/1758-2946-3-33

OECD. Test No. 428: Skin Absorption: In Vitro Method: OECD Publishing.

Plate, T. (2016). track: Store Objects on Disk Automatically. R package version 1.1.9. https://CRAN.R-project.org/package=track

Plummer, M. (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. https://CRAN.R-project.org/package=rjags

Plummer, M., Best, N., Cowles, K. & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC, R News, vol 6, 7-11

Potts, R. O., & Guy, R. H. (1992). Predicting Skin Permeability. Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists, 9(5), 663-669. doi:10.1023/A:1015810312465

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riviere, J. E., & Brooks, J. D. (2007). Prediction of dermal absorption from complex chemical mixtures: incorporation of vehicle effects and interactions into a QSPR framework. SAR QSAR Environ Res, 18(1-2), 31-44. doi:10.1080/10629360601033598

Russell, L. M., Guy, R. H., 2009. Measurement and prediction of the rate and extent of drug delivery into and through the skin. Expert Opinion on Drug Delivery. 6, 355-369.

Shah JC, Kaka I, Tenjarla S, Lau SWJ, Chow D (1994) Analysis of percutaneous permeation data: II. Evaluation of the lag time method. INT J PHARM 109(3):283-290 doi:10.1016/0378-5173(94)90390-5

Su Y-S. Yajima, M. (2015). R2jags: Using R to Run 'JAGS'. R package version 0.5-7. https://CRAN.R-project.org/package=R2jags

Viswanadhan, V. N., Ghose, A. K., Revankar, G. R., & Robins, R. K. (1989). Atomic Physicochemical Parameters for 3 Dimensional Structure Directed Quantitative Structure - Activity Relationships .4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally-Occurring Nucleoside Antibiotics. Journal of Chemical Information and Computer Sciences, 29(3), 163-172. doi:DOI 10.1021/ci00063a006

Wang, R. X., Fu, Y., & Lai, L. H. (1997). A new atom-additive method for calculating partition coefficients. Journal of Chemical Information and Computer Sciences, 37(3), 615-621. doi:DOI 10.1021/ci960169p

Wickham, H. & Bryan, J. (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Wickham, H. (2017). scales: Scale Functions for Visualization. R package version 0.5.0. https://CRAN.R-project.org/package=scales

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Xiao P, Imhof RE (1997) Optothermal measurement of stratum corneum thickness and hydration-depth profile. SPIE Proc 2970:276-286

Yalkowsky, S. H. and Valvani, S. C. (1980), Solubility and partitioning I: Solubility of nonelectrolytes in water. J. Pharm. Sci., 69: 912-922. doi:10.1002/jps.2600690814

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Abbreviations/Glossary

Term	Explanation
%DA	percent dermal absorption
(Q)SAR	(quantitative) structure activity relationship
3D	three-dimensional
a.s.	active substance
AI	experimental solution of active substance in solvent
ANN	artificial neural network
H_a *or* B:	hydrogen bond acceptor activity (summation) or basicity
BfR	Bundesinstitut für Risikobewertung
block aggregates	averages between replicates of one experiment
BP	boiling point
BPI	BENAKI Phytopathological Institute
BREM	Bayesian random effects model
c concentrate	experimental concentration of the active substance
DAME	Dermal Absorption Model for Extrapolation
DAR	Draft Assessment Report
DU	Durham University
e.g.	for example
EFSA	European Food Safety Authority
EHOMO	energy of the highest occupied molecular orbital
ELUMO	energy of the lowest unoccupied molecular orbital
EPI Suite	Estimation Programs Interface Suite
Eq.	equation
HA, HBA	number of Hydrogen bond acceptor sites
HB	Sum of Hydrogen bond acceptors and donors
HD, HBD	number of Hydrogen bond donor sites
HOMO	highest occupied molecular orbital
IJC	Instant JChem
J_{max},	maximum flux
Kow	octanol/water partition coefficient
Kow, K(OW)	octanol-water partition coefficient
k_p	permeability coefficient or permeation coefficient
log Soct	logarithm (base 10) of the solubility in octanol
LogP, LogPow, LogKow	logarithm (base 10) of the octanol/water partition coefficient
logSaq	logarithm (base 10) of the water solubility
LUMO	lowest unoccupied molecular orbital
M1	model 1 - Frasch 2002
M10	model 10 – Riviere Brooks 2007 No. 2
M11	model 11 – Atobe 2015
M2	model 2 – Potts & Guy 1992
M3	model 3 – Magnusson 2004
M4	model 4 – Fu 2002 Algorithmic solution
M5	model 5 – Fu 2002 Artificial Neural Network
M6	model 6 – Milewski 2012
M7	model 7 – Dancik / NIOSH

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Term	Explanation
M8	model 8 – Potts & Guy 1995
M9	model 9 – Riviere Brooks 2007 No. 1
Mdon,0	initial mass of substance present in donor fluid
MF	mixture factor
mixture models	models that estimate individual compound (in this case pesticide a.s.) penetration by additional consideration of effects of the co-formulants/vehicle, expressed as Mixture Factor (MF). The results obtained by these models do not reflect the absorption of each single component in the "mixture"
mLogKow	mixture logarithm (base 10) of the octanol/water partition coefficient; sum of the weight percentage for each of the component´s logarithm (base 10) of the octanol/water partition coefficient of a mixture
MM	mixture model (for explanation of the term "mixture model" see above)
MP	melting point
MPS	Model parameter score: the n(th)-root of the product of the parameter scores of all parameters associated with a model
MR	molar refractivity
MRL	Multiple linear regression model
mTpSA	mixture topological polar surface area; sum of the weight percentage for each of the component´s topological polar surface area of a mixture
MV	molecular volume
MW	molecular weight
NA	not available
NIOSH	National Institute for Occupational Safety and Health
No.	number
NS	not specified
pKa	negative logarithm (base 10) of the acid dissociation constant
Qh,	sum of the net atomic charges of the hydrogen atoms bound to nitrogen or oxygen atoms
QO.N,	the sum of the absolute values of the net atomic charges of oxygen and nitrogen atoms which are hydrogen-bond acceptors
RAR	Renewal Assessment Report
replicate	Corresponds to information regarding one Franz cell in the *in vitro* experiment. The absorption of a substance at a given concentration and at givebn experimental conditions is derived by averaging the replicates that belong to the same treatment group.
RMSE	root-mean-square error
RR	Registration Report
S	Abraham descriptor: dipolarity/polarity of the solute
Saq	water solubility
SC	*stratum corneum*
SSM	single substance model
t	exposure time
tlag, t_{lag}	lag time
TNO	TNO innovation for life

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Term	Explanation
TpSA	topological polar surface area
V	Abraham descriptor: solute´s McGowan characteristic volume
Vdon	volume of donor fluid
Ve	Van der Waals volume
VE	viable epidermis
VP	vapour pressure
Vsc	volume of *stratum corneum*
WP	work package

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix A – Literature Search

A search protocol was drafted and, following internal review, presented to EFSA for refinement at a meeting in Parma on June 16, 2017. The draft search protocol was adapted in conformity with the suggestions received during that meeting. Amongst others, it was tested whether relevant papers would be lost when not including "Environmental Science" as a discipline in the Scopus search. As this turned out to be the case, the discipline "Environmental Science" was not excluded. For further validation, it was examined whether relevant papers published after 1989 as listed in table 1-1 of Buist (2016) would be obtained using the revised search strategy in Scopus (see Table A.1). Of the 27 papers 15 were found, 12 were missed and 4 were not listed in Scopus. Based on this outcome, the Scopus search protocol was refined and the results checked again against the same list of papers. This was done in an iterative fashion, until satisfactory results were obtained. Finally, 11 of the 12 papers missed in the original search were retrieved in the final search.

Table A.1: Verification of search results in Scopus.

Reference	Hit in Scopus		Reference	Hit in Scopus	
	Originally	Finally		Originally	Finally
Abraham and Martins (2004)	Yes	Yes	McKone and Howd (1992)	No	No
Barratt (1995)	No	Yes	Mitragotri (2002)	Yes	Yes
Buchwald and Bodor (2001)	No	Yes	Moody and MacPherson (2003)	Yes	Yes
Cleek and Bunge (1993)	Yes	Yes	Moss and Cronin (2002)	Yes	Yes
Dearden *et al.* (2000)	No	Yes	Patel *et al.* (2002)	Yes	Yes
Fiserova-Bergerova *et al.* (1990)	No	Yes	Potts and Guy (1992)	No	Yes
Flynn (1990)	No (not in Scopus)	n/a	Potts and Guy (1995)	Yes	Yes
Fujiwara *et al.* (2003)	No	Yes	Pugh *et al.* (2000)	Yes	Yes
Geinoz *et al.* (2004)	Yes	Yes	Tayar *et al.* (1991)	Yes	Yes
Guy and Potts (1992)	No	Yes	ten Berge (2009)	Yes	Yes
Hostynek and Magee (1997)	Yes	Yes	USEPA (2004)	No (not in Scopus)	n/a
Lien and Gao (1995)	Yes	Yes	Vecchia and Bunge (2002)	No (not in Scopus)	n/a
Magee (1998)	No (not in Scopus)	n/a	Wilschut *et al.* (1995)	Yes	Yes
Magnusson *et al.* (2004)	Yes	Yes			

The Scopus search protocol was then translated into PubMed and ToxLine search protocols, taking into account the recommendations made at the meeting with EFSA in Parma on June 16, 2017.

Scopus search

Below the Scopus search string is listed. It was executed on June 20, 2017, and resulted in 1,430 hits.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

(((TITLE-ABS-KEY (skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key(skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key(skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal penetration") AND pubyear > 1989) OR (title-abs-key("dermal permeation") AND pubyear > 1989)) AND (qsar)) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key (skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key(skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal penetration") AND pubyear > 1989) OR (title-abs-key("dermal permeation") AND pubyear > 1989)) AND ("in silico")) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key (skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key(skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal penetration") AND pubyear > 1989) OR (title-abs-key("dermal permeation") AND pubyear > 1989)) AND (qspr)) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key (skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key(skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal penetration") AND pubyear > 1989) OR (title-abs-key("dermal permeation") AND pubyear > 1989)) AND ("mathematical model")) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key(skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key(skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal penetration") AND pubyear > 1989) OR (title-abs-key("dermal permeation") AND pubyear > 1989)) AND ("artificial neural network")) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key(skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key (skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND pubyear > 1989) OR (title-abs-key("dermal penetration") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal permeation") AND pubyear > 1989)) AND (algorithm)) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key(skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key (skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND pubyear > 1989) OR (title-abs-key("dermal penetration") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal permeation") AND pubyear > 1989)) AND ("machine learning")) OR (((title-abs-key(skin W/10 absorption) AND pubyear > 1989) OR (title-abs-key(skin W/10 penetrat*) AND pubyear > 1989) OR (title-abs-key (skin W/10 permeat*) AND pubyear > 1989) OR (title-abs-key("dermal absorption") AND pubyear > 1989) OR (title-abs-key("dermal penetration") AND PUBYEAR > 1989) OR (TITLE-ABS-KEY("dermal permeation") AND pubyear > 1989)) AND ("statistical model")) AND (limit-to (language , "English") OR limit- to (language , " German")) AND (limit-to (subjarea , "PHAR") OR limit- TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "BIOC") OR LIMIT-TO (SUBJAREA , "CHEM") OR LIMIT-TO (SUBJAREA, "ENVI") OR LIMIT-TO (SUBJAREA , "AGRI") OR LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA , "VETE") OR LIMIT-TO (SUBJAREA , "MATH") OR LIMIT-TO (SUBJAREA, "MULT') OR LIMIT-TO (SUBJAREA , "Undefined")) AND (exclude(doctype , "cp") OR exclude(doctype , "sh") OR EXCLUDE (DOCTYPE , "le") OR EXCLUDE (DOCTYPE , "ed") OR EXCLUDE (DOCTYPE , "no")) AND (EXCLUDE (SUBJAREA , "BUSI") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA, "DENT')| OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA, "ENER") OR EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA, "PHYS") OR EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA, "HEAL") OR EXCLUDE (SUBJAREA , "IMMU") OR EXCLUDE (SUBJAREA, "SOCI") OR EXCLUDE (SUBJAREA , "NURS") OR EXCLUDE (SUBJAREA , "PSYC"))

PubMed

Below the PubMed search string is listed. It was executed on June 20, 2017, and resulted in 1,416 hits.

((((((qsar OR "in silico" OR qspr OR "mathematical model" OR "artificial neural network" OR "algorithm" OR "machine learning" OR "statistical model" OR predict*[MeSH Terms])) OR (qsar[Title/Abstract] OR "in silico"[Title/Abstract] OR qspr[Title/Abstract] OR "mathematical model"[Title/Abstract] OR "artificial neural network"[Title/Abstract] OR "algorithm"[Title/Abstract] OR

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

"machine learning"[Title/Abstract] OR "statistical model"[Title/Abstract] OR predict*[Title/Abstract])) AND ((((skin AND absorption) OR (skin AND penetrat*) OR (skin AND permeat*) OR "dermal absorption" OR "dermal penetration" OR "dermal permeation"[MeSH Terms])) OR ((skin AND absorption) OR (skin AND penetrat*) OR (skin AND permeat*) OR "dermal absorption" OR "dermal penetration" OR "dermal permeation"[MeSH Terms]))) AND ("german"[Language] OR "english"[Language])) AND ("1990/01/01"[Date - Publication] : "3000"[Date - Publication])

ToxLine

Below the ToxLine search string is listed. It was executed on June 20, 2017, and resulted in 246 hits.

((qsar OR "in silico" OR qspr OR "mathematical model" OR "artificial neural network" OR "algorithm" OR "machine learning" OR "statistical model" OR predict*) AND ((skin AND absorption) OR (skin AND penetrat*) OR (skin AND permeat*) OR "dermal absorption" OR "dermal penetration" OR "dermal permeation")) AND 1990:2017 [yr] AND (eng [la] OR ger [la]) AND NOT PubMed [org] AND NOT pubdart [org]

Effectivity check

For verification of the effectivity of the literature searches, BfR and TNO analysed whether all relevant papers relating to mathematical models for dermal absorption as cited Buist (2016) (Table A.1) or in the recently updated EFSA Guidance on Dermal Absorption (EFSA, 2017) were found. The four papers that were not listed in Scopus were also absent from PubMed and ToxLine and could therefore not be retrieved by any of the searches. The McKone and Howd paper referred to in Buist (2016), which was not retrieved by the Scopus search as described above, was also not present among the hits of the searches in PubMed and ToxLine. The result of the effectivity check against the papers cited in the EFSA Guidance is presented in Table A.2.

Table A.2: Results of effectivity check with papers cited in the EFSA dermal absorption guidance (EFSA, 2017).

Paper	Hit?
Anissimov *et al.* (2013)	Yes
Buist (2016)	No (not present in databases)
Buist *et al.* (2010)	Yes
Gute *et al.* (1999)	Yes
Mitragotri *et al.* (2011)	No
Russell and Guy (2009)	Yes

Analysis of the key words associated with the missed papers showed that the likely cause of missing them was the absence of the term "permeability" from the search terms. Therefore additional searches were performed in the three literature databases to amend this omission. The results are reported below, and the amended literature search now includes the two originally missed papers (McKone and Howd, 1992; Mitragotri *et al.*, 2011).

Additional Scopus search

Below the Scopus search string for the additional literature search is provided. The search was executed on July 6, 2017, and resulted in 209 hits. After unduplication in the LRT, 104 new unique records remained.

((TITLE-ABS-KEY ("skin permeability") AND pubyear > 1989) OR (title-abs-key ("dermal permeability") AND pubyear > 1989)) AND ((TITLE-ABS-KEY ("mathematical model") AND pubyear > 1989) OR (TITLE-ABS-KEY ("artificial neural network") AND pubyear > 1989) OR (TITLE-ABS-KEY ("machine learning") AND pubyear > 1989) OR (TITLE-ABS-KEY ("statistical model") AND pubyear > 1989) OR (TITLE-ABS-KEY (algorithm) AND pubyear > 1989)) AND (limit-to (language , "English") OR limit- to (language , " German")) AND (limit-to (subjarea , "PHAR") OR limit- TO (SUBJAREA , "MEDI") OR

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

LIMIT-TO (SUBJAREA , "BIOC") OR LIMIT-TO (SUBJAREA , "CHEM") OR LIMIT-TO (SUBJAREA, "ENVI") OR LIMIT-TO (SUBJAREA , "AGRI") OR LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA , "VETE") OR LIMIT-TO (SUBJAREA , "MATH") OR LIMIT-TO (SUBJAREA, "MULT') OR LIMIT-TO (SUBJAREA , "Undefined")) AND (exclude(doctype , "cp") OR exclude(doctype , "sh") OR EXCLUDE (DOCTYPE , "le") OR EXCLUDE (DOCTYPE , "ed") OR EXCLUDE (DOCTYPE , "no")) AND (EXCLUDE (SUBJAREA , "BUSI") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA, "DENT')| OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA, "ENER") OR EXCLUDE (SUBJAREA , "ENGI") OR EXCLUDE (SUBJAREA, "PHYS") OR EXCLUDE (SUBJAREA , "MATE") OR EXCLUDE (SUBJAREA, "HEAL") OR EXCLUDE (SUBJAREA , "IMMU") OR EXCLUDE (SUBJAREA, "SOCI") OR EXCLUDE (SUBJAREA , "NURS") OR EXCLUDE (SUBJAREA , "PSYC"))

Additional PubMed search

Below the PubMed search string for the additional literature search is provided. The search was executed on July 6, 2017, and resulted in 181 hits. After unduplication in the LRT, to which the result from Scopus already had been added, 29 new unique records remained.

((((("quantitative structure-activity relationship"[MeSH Terms] OR ("quantitative"[All Fields] AND "structure-activity"[All Fields] AND "relationship"[All Fields]) OR "quantitative structure-activity relationship"[All Fields] OR "qsar"[All Fields]) OR "in silico"[All Fields] OR qspr[All Fields] OR "mathematical model"[All Fields] OR "artificial neural network"[All Fields] OR "algorithm"[All Fields] OR "machine learning"[All Fields] OR "statistical model"[All Fields]) AND ((("skin"[MeSH Terms] OR "skin"[All Fields]) AND ("permeability"[MeSH Terms] OR "permeability"[All Fields])) OR (dermal[All Fields] AND ("permeability"[MeSH Terms] OR "permeability"[All Fields])))) AND ("german"[Language] OR "english"[Language])) AND ("1990"[PDAT] : "3000"[PDAT])

Additional ToxLine search

Below the ToxLine search string for the additional literature search is provided. The search was executed on July 6, 2017, and resulted in 27 hits. After unduplication in the LRT, to which the result from Scopus and PubMed already had been added, 7 new unique records remained.

((((qsar OR "in silico" OR qspr OR "mathematical model" OR "artificial neural network" OR "algorithm" OR "machine learning" OR "statistical model") AND ((skin AND permeability) OR (dermal permeability)))) AND (eng [la] OR ger [la])) AND 1990:2017 [yr] AND NOT PubMed [org] AND NOT pubdart [org]

Google search

In addition to searches in the literature databases, an internet search using Google was performed. Below the Google search string is provided.

(qsar OR "in silico" OR qspr OR "mathematical model" OR "artificial neural network" OR "algorithm" OR "machine learning" OR "statistical model" OR predict*) AND ((skin AND absorption) OR (skin AND penetrat*) OR (skin AND permeat*) OR "dermal absorption" OR "dermal penetration" OR "dermal permeation" OR ("skin permeability") OR ("dermal permeability"))

It was implemented on July 6, 2017 and resulted in seven webpages of hits, of which three were not peer-reviewed papers and not considered relevant based on title and linked webpage. The detailed results and their relevance assessment are reported in Table A.3. Three new references were included in the LRT.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table A.3: Google search results. 1 Jan 2016 – 6 Jul 2017, Sorted by relevance, all results, clear, about 11.200 results (0,58 seconds), seven pages displayed, first three pages copied.

Predicting the Rate of Skin Penetration Using an Aggregated ... - NCBI https://www.ncbi.nlm.nih.gov/pubmed/28335598 by M Lindh - 2017 Apr 17, 2017 - Predicting the Rate of Skin Penetration Using an Aggregated Conformal ... models use calculated descriptors and can quickly predict the skin permeation rate of ...	Publication
Skin models for the testing of transdermal drugs - NCBI - NIH https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5076797/ by E Abd - 2016 - Cited by 1 - Related articles Oct 19, 2016 - More commonly used models to conduct skin-permeation studies are ex vivo .... maximum flux and a trend toward decreased dermal penetration rates.35 .... in the interpretation of dermal absorption for human risk assessment, to predict ...	Publication
Simple Predictive Models of Passive Membrane Permeability ... pubs.acs.org/doi/pdf/10.1021/acs.jcim.6b00005 by SSF Leung - 2016 - Cited by 7 - Related articles May 2, 2016 - biological processes, such as intestinal absorption, skin penetration, or blood-brain barrier permeation, involve permeation of molecules across ... for drug-likeness described by Lipinski's rule-of-five and QSPR models for permeability.5.	Publication
Data-based modeling of drug penetration relates human skin barrier ... www.pnas.org/content/114/14/3631.full.pdf by R Schulz - 2017 Mar 20, 2017 - Thereby we can predict short-time drug penetration, where experimental ... eral models for the permeation of drugs through skin exist, which incorporate the skin ...	Publication
Surging footprints of mathematical modeling for prediction of ... www.sciencedirect.com/science/article/pii/S1818087617301344 by N Goyal - 2017 Feb 23, 2017 - Mathematical models of skin permeability are highly relevant with respect to ... Themathematical model is based on the hypotheses that consider ..... model (concentration in model membrane is related to the penetration depth) and ...	Publication
Skin Permeation Rate as a Function of Chemical Structure (PDF ... https://www.researchgate.net/.../7059484_Skin_Permeation_Rate_as_a_Function_of_... Jun 4, 2017 - Multilinear and nonlinear QSAR models were built for the skin permeation rate (Log K(p)) of ... penetration of other substances by perturbing the barrier function.	Publication
Carbon nanotube membranes to predict skin permeability of compounds www.tandfonline.com/doi/full/10.1080/10837450.2016.1221430 by S Ilbasmis-Tamer - 2017 - Cited by 2 - Related articles Aug 30, 2016 - Abstract. In the present study, carbon nanotube (CNT) membranes were prepared topredict skin penetration properties of compounds. A series of penetration ...	Publication
[PDF]Alkylglycerol Derivatives, a New Class of Skin Penetration Modulators www.mdpi.com/1420-3049/22/1/185/pdf by SA Bernal-Chávez - 2017 Jan 22, 2017 - as well as in vitro skin permeation studies, were performed in order to ... [9] showed, in an in silico study, that 1-O-alkylglycerols, particularly 1-O-octadecyl.	Publication
[PDF]COMPUTER SIMULATION OF SKIN ABSORPTION AND ... - Xemet https://www.xemet.com/media/filer.../45/.../skin_simulation_brochure_10-15.pdf Apr 17, 2016 - Benefits of computer simulation of skin absorption/permeation ... Simulation also enables evaluation of vehicle effects, where the vehicle (penetration enhancer) is ... The underlyingmathematical model describes simultaneous absorption of ...	Commercial in silico model (presentation saved)
Insights into skin permeation: from theory to practice - CECAM https://www.cecam.org/workshop-1418.html	Workshop: Insights into skin permeation:

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

by C Das Jan 3, 2017 - The topmost layer of the skin, the stratum corneum[1], comprises rigid non-viable ... include iontophoresis (use of an electrical potential to enhance penetration), ... Although in-silicomodels are widely used but none of the available models are ...	from theory to practice. Not relevant
Page 2 of about 11.200 results (0,64 seconds) Search Results Impact of semi-solid formulations on skin penetration of iron oxide ... https://jnanobiotechnology.biomedcentral.com/articles/10.1186/s12951-017-0249-6 by UM Musazzi - 2017 Feb 17, 2017 - For these reasons, the **skin penetration** of IONs has been gaining increasing ... of the nanoparticles are not the only critical determinants for their **skin permeation.** ..... Use of in vitro human**skin** membranes to model and **predict** the effect of ...	. Publication
Using Neural Networks and Ensemble Techniques based on Decision ... link.springer.com/chapter/10.1007/978-981-10-4166-2_8 by E Bušatlić - 2017 Mar 14, 2017 - This paper presents results of development of **Artificial Neural Network** (ANN) for prediction of **skin permeability.** The performance of developed ANN was ...	Book
Skin penetration,Skin permeation IVPT,Solubility studies,Emulsion www.tenguspharma.com/skin-penetration-studies/ Mar 3, 2017 - Tergus is the recognized leader in the **skin permeation and penetration** studies. ... formulation to competing products and help **predict** in vivo **penetration**.	Company, in vitro predictions
[PDF]Prediction of Dermal Permeability Coefficient of Nevirapine—Effect of ... https://file.scirp.org/pdf/PP_2016072816082094.pdf by CJ Mbah - 2016 - Related articles Jul 25, 2016 - As the rate of **penetration** into the **skin** is quantitatively ... cate that **permeability coefficient** can be a more reliable parameter to **predict** transdermal ab- sorption of ...	Publication
[PDF]SkinCourse in Saarbrücken March2017 - Universität des Saarlandes www.uni-saarland.de/fileadmin/user_upload/.../fr82.../Flyer_SkinCourse_online.pdf Mar 13, 2017 - **Permeation model**: Franz **diffusion cells**. **Penetration** model: Saarbrücken model &**skin** segmentation. Confocal RAMAN microscopy for **dermal** analysis. **In silico** ...	University course
Generic Drug User Fee Amendments of 2012 > FY2016 Regulatory ... https://www.fda.gov/ForIndustry/UserFees/GenericDrugUserFees/ucm549169.htm Mar 29, 2017 - Figure 2: Dermal pharmacokinetics of acyclovir in vivo in human subjects ... the **skin**from a topical product, and then **permeate through** the layers of the **skin** and ..... Sinner F. Can we**predict skin penetration** by using TEWL or **skin** impedance?	Regulatory information
Effect of Size and Surface Charge of Gold Nanoparticles on their Skin ... https://www.nature.com › Scientific Reports › Articles Mar 28, 2017 - However, the mechanisms by which these AuNPs **penetrate** are not well understood. ... In contrast, Liu *et al.* showed that viable human **skin** resists **permeation of** small ..... Help Reduce Drug Candidate Attrition and Move Us Beyond **QSPR**?	Publication
Drug permeation and barrier damage in ... - Oxford Academic https://academic.oup.com/jac/article-pdf/71/6/1578/11282328/dkw012.pdf by K Van Bocxlaer - 2016 - Cited by 3 - Related articles Feb 21, 2016 - Drug **permeation and** barrier damage in Leishmania-infected mouse **skin** ... paromomycin, for Leishmania-infected **skin** compared with uninfected **skin**. ..... Altered **penetration** of polyethyl- .... Studies: In **Skin**, In Vitro,	Publication

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

and **In Silico** Models.	
The EDETOX Database; EDETOX; Newcastle University https://research.ncl.ac.uk › EDETOX Jul 15, 2016 - Development of **dermal absorption** database for cosmetic chemicals from existing ... on in vitro human **skin permeation-** Data from updated EDETOX database. ... JJ (2008) Improving the applicability of (Q)SARs for percutaneous **penetration** in ... decision tree to **predict** dermally delivered systemic dose for comparison with ...	In vitro and in vivo study results
PREDICTING THE ABSORPTION RATE OF CHEMICALS THROUGH ... https://uhra.herts.ac.uk/.../12239113%20Ashrafi%20Parivash%20final%20PhD%20sub... by P Ashrafi - 2016 - Cited by 1 - Related articles May 15, 2016 - I attempt to find optimal values of **skin permeability using** GP optimisation algorithms within small datasets. ... 2 **Skin Permeability and** the Traditional **QSAR**/**QSPR** Approaches. 7 ..... ence on Perspectives in Percutaneous **Penetration**. France ...	Thesis check whether also published in paper
Dominik Selzer - Google Scholar Citations scholar.google.com/citations?user=-k7lUIkAAAAJ&hl=en Saarland University; Scientific Consilience GmbH - scientific-consilience.com Apr 1, 2017 - Towards drug quantification in human skin with confocal Raman microscopy ... A strategy for in-silico prediction of skin absorption in man ... Human Native and Reconstructed SkinPreparations for In Vitro Penetration and Permeation Studies.	Literature list
Evaluation of skin absorption of drugs from topical and transdermal ... www.scielo.br/pdf/bjps/v52n3/2175-9790-bjps-52-03-00527.pdf by ALM Ruela - 2016 - Related articles Sep 15, 2016 - of skin penetration and permeation of drugs from topical and transdermal ..... transdermal flux of the drugs to predict their in vivo behavior from different drug ...	Publication
Colloidal Dispersions (Liposomes and Ethosomes) for Skin Drug ... www.asiapharmaceutics.info/index.php/ajp/article/download/728/467 by M Nimmathota - 2016 - Related articles Sep 7, 2016 - permeate or penetrate through the skin. ... Routes of drug penetration through theskin[10-12] .... predict the permeation profiles of liposomal vesicles through.	Publication
[PDF]Molecular dynamics simulation study of translocation of fullerene C 60 ... pubs.rsc.org/-/content/articlepdf/2017/nr/c6nr09186e by R Gupta - 2017 Feb 22, 2017 - fullerene-based peptide penetrated intact skin, and mechan- ical stressors, such as those ... permeation of pristine fullerene C60 with a model skin mem- brane using ..... candidate attrition and move us beyond QSPR, Chem. Biol. Drug Des.	Publication
European Commission : CORDIS : Projects & Results Service : Final ... cordis.europa.eu/result/rcn/182206_en.html May 17, 2016 - Final Report Summary - COSMOS (Integrated In Silico Models for the ..... In addition,skin permeability data were donated from the EDETOX Database and ..... These emphasise that estimation of skin penetration may be challenging for ...	Check for in silico model (COSMOS project)
[PDF]In Silico Estimation of Skin Concentration of Dermally ... - Peertechz https://www.peertechz.com/.../in-silico-estimation-of-skin-concentration-of-dermally-... Jan 23, 2017 - were determined from the in vitro permeation data through full-thickness skin and stripped skin after application of ... chemical; In silico estimation; Skin permeation; Skin ..... In vivo methods for the analysis of the penetration of topically applied.	Publication
In Vitro Testing & Other Services | AltTox.org alttox.org › Resource Center Jun 14, 2017 - Absorption Systems, LP (Exton, Pennsylvania, U.S.): preclinical contract ...	In vitro

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

UK):dermal research and development including in vitro human skin permeation .... in silico, the in vitro epidermis screening test (INVEST) and other ex vivo skin models. ..... including GLP services, for skinirritation and corrosion, dermal penetration, ...	
transdermal penetration of acyclovir in the presence and absence.of ... https://dspace.nwu.ac.za/bitstream/handle/10394/301/myburgh_m.pdf?sequence=1 by M Myburgh - 2003 Jan 4, 2016 - acyclovir is the permeation in sufficient amounts to deeper layers of the skin and ... acyclovir, transdermal delivery, permeation, penetration enhancers, terpenes ..... To predict skintransport of a drug molecule, it is important to understand the ...	In vitro
Prediction of human pharmacokinetic profile after transdermal drug ... jpharmsci.org/article/S0022-3549(17)30159-4/pdf by S Yamamoto Mar 2, 2017 - compartment model using in vitro human skin permeation parameters as zero-orderabsorption ... represent the stratum corneum and viable skin may predict in vivo human ... skin penetration of drug substances have been proposed.14-16.	Publication
[PDF]Investigation of Caffeine Permeation in the Skin PAMPA Model abstracts.aaps.org/Verify/AAPS2016/PosterSubmissions/26W0400.pdf Nov 7, 2016 - Caffeine was selected as a model permeant as the skin permeation properties of ...penetration enhancers for caffeine were selected, namely dimethyl ...	In vitro

Checked additional 4 pages (of seven in total): no relevant hits.

Unduplicating in EndNote

The records resulting from the searches in Scopus, PubMed and ToxLine were exported to EndNote. Since the automatic unduplication function of EndNote is limited to records directly imported via a web search executed in EndNote, this was performed manually. Furthermore, all references to the Reporter Database of NIH were removed as this database represents a repository of (future) projects rather than of scientific reports or papers. This removal and unduplication resulted in a final list of a total of 2533 records (2111 from the main search) to be exported to the TNO Literature Review Tool (LRT). This was more than twice the limit intended in Technical Offer. However, since attempts to further cut down this number to 1000 would increase the risk of missing relevant papers, all were taken forward to the next phase in the selection procedure. The EndNote file is available upon request at EFSA.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix B – Literature Relevance Assessment

Relevance check on title and abstract in the TNO Literature Review Tool (LRT)

Main search

When the 2111 records from the main search as described in 1.2 to 1.4 were imported from EndNote into the LRT, 2069 unique records remained, which all received an unique five-digit identifier, the LRT-ID. These publications were first checked for relevance based on their title. In the context of this check, "relevant papers" were defined as "papers addressing mathematical models of dermal absorption". Records considered irrelevant based on their title were not taken forward to the evaluation according to scientific criteria (see chapter 2). The remaining publications were further examined for relevance based on their abstract. However, in some cases abstracts were not (publically) available at the time of the execution of the relevance check. These papers were not excluded due to missing abstract at this initial step of relevance check. The relevance of these papers according to the abstract was evaluated, when the abstract was made available after ordering full-texts via the library service. Records considered irrelevant based on their abstract were also excluded form the next selection phase. The reasons for exclusion were documented. The following questions were answered in the tool in order to come to this selection to should be answered with YES or NO (with a number of sub-options for the latter):

1. Relevant based on title?

- ☐ **Yes**
- ☐ **No** - No mathematical model of dermal absorption
- ☐ **No** - Other

2. Relevant based on abstract?

- ☐ **Yes**
- ☐ **No** - No mathematical model of dermal absorption
- ☐ **No** - Refers to models, but does not describe them
- ☐ **No** - PBPK model for a single substance
- ☐ **No** - Mixed model with in silico and experimental input
- ☐ **No** - Other

Only papers for which both questions were answered with YES will be considered in the scientific review. Following internal peer review of this initial relevance assessment (see section below), 266 papers were selected for the scientific review. Results of the relevance assessment for all individual references are available in an MS Excel file upon request at EFSA.

Quality control relevance check

Internal quality control was performed regarding the relevance check. Approximately 5% of all publications retrieved during the main literature search, i.e. approx. 100 papers, were evaluated for concordance by the internal reviewer (Nadine Engel, BfR). Each twentieth paper in the list of publications (starting with number 1, then 21, then 41, then 61, etc.) were blindly selected for this procedure. When the reviewer did not concord, the reason of deviation was be registered. It was defined prior to starting the assessment that if the percentage of unjustly EXCLUDED papers is higher than 5% of the papers checked (i.e. more than 5 papers), an additional 10% of the scored papers will be checked by selecting and checking every tenth paper, starting with number 5, then numbers 15, 25, 35, etc.). If the percentage of unjustly EXCLUDED papers remains higher than 5% (i.e. more than approx. 10 papers), all remaining scored papers would be checked.

Out of the 104 publications checked blindly, the internal reviewer did not agree with the conclusions of the assessor in 4 cases. In one of these cases the assessor did not agree with the conclusions of

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

the internal reviewer. After discussion, the internal reviewer accepted the evaluation of the assessor. Therefore, in 3 out 104 cases the first assessment did not pass internal QC, which amounts to 2.9% and was regarded as acceptable according to the established protocol.

Additional searches

Of the 143 records resulting from the additional searches (see Appendix A for background), 33 publications were considered to be relevant based on title and/or abstract (when present). Of these 11 were identified as duplicates of papers already found in the main search. Accordingly, 22 relevant papers resulting from the additional search were added to the 266 publications from the main search yielding a total 288 relevant papers. These were subjected to the scientific appraisal as described in chapter 2.2.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix C – Methodology for Stage 2 of the Systematic Literature Appraisal

The second stage of the scientific review was executed in Excel. The issues addressed in the second stage of the scientific review are listed in Table C.1 including the questions and the selectable answers, respectively. Models for which questions 1, 2, or 5 were answered with "No", or question 3 was answered with "Yes", were not further considered and the questions that follow thereafter were not answered for those models.

Table C.1: Questions answered during the second stage of the scientific review.

Question 1: Algorithm based model?		
Answers	Yes/No	
Options if "Yes"		Well-defined and specific algorithm
		Not a well-defined and specific algorithm
Options if "No"		Reliable reconstruction possible
		No reliable reconstruction possible
Motivation		<free text>
Question 2: Addresses a clearly defined need in pesticide RA?		
Answer	Yes/No	
Options if "Yes"		Global model predicting k_D
		Global model predicting J_{max}
		Global model predicting % absorption
		Local model predicting k_D for pesticide group(s)
		Local model predicting J_{max} for pesticide group(s)
		Local model predicting % absorption for pesticide group(s)
		Other
Specify pesticide group(s) if local model		<free text>
Specify "Other" if selected		<free text>
Options if "No"		does not predict a dermal absorption parameter
		Local model predicting k_D for non-pesticide group(s)
		Local model predicting J_{max} for non-pesticide group(s)
		Local model predicting % absorption for non-pesticide group(s)
		categorical model of dermal absorption
		Other
Specify non-pesticide group(s) if local model		<free text>
Specify "Other" if selected		<free text>
Question 3: Are there more recent similar models by the same research group?		
Answer	Yes/No	<if yes, specify paper>
Question 4: Details of the training set provided?		
Answer	Yes/No	
Options if "Yes"		Chemical IDs of all members of the training set
		Measured absorption parameters
		Parameter values used in model
		Other
specify "Other" if selected		<free text>
Question 5: Can the model be reproduced based the materials and methods of the paper?		
Answer	Yes/No	
Options if "Yes"		Completely
		Partly
Motivation		<free text>
Question 6: Internal validation performed?		
Answer	Yes/No	

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Options if "Yes"		**Parameter**	**Value**
		R2	<number>
		Adjusted R2	<number>
		RMSE	<number>
		number of predictors	<number>
		number in training set	<number>
		Other	<number>
specify "Other" if selected		<free text>	
Question 7: External validation performed?			
Answer	Yes/No		
Options if "Yes"		**Parameter**	**Value**
		R2	<number>
		Adjusted R2	<number>
		RMSE	<number>
		number of predictors	<number>
		number in validation set	<number>
		Other	<number>
specify "Other" if selected		<free text>	
Question 8: Can the model be interpreted mechanistically?			
Options	A	Clear how input parameters might cause changes in absorption rate and model does specify quantitative influence of parameter values on predicted absorption parameter	
	B	Clear how input parameters might cause changes in absorption rate, but model does NOT specify quantitative influence of parameter values on predicted absorption parameter	
	C	NOT clear how input parameters might cause changes in absorption rate, but model does specify quantitative influence of parameter values on predicted absorption parameter	
	D	NOT clear how input parameters might cause changes in absorption rate and model does NOT specify quantitative influence of parameter values on predicted absorption parameter	
Motivation	<free text>		
Question 9: Proprietary model?			
Answer	Yes/No		

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix D – Detailed Results of the Systematic Literature Appraisal

Table D.1: Algorithm models possibly relevant for pesticide risk assessment - Single substance models.

Citation	Model No.	Model algorithm(s)	LRT-ID
Abraham M. H. and Martins F. (2004)	1	$\log k_p = 0.106R2 - 0.473\pi H2 - 0.473\Sigma\alpha H2 - 3.00\Sigma\beta H2 + 2.296Vx - 1.866$	26295
Abraham M. H. and Martins F. and Mitchell R. C. (1997)	1	$\log k_p = 0.44R2 - 0.49\pi H2 - 1.48\Sigma\alpha H2 - 3.44\Sigma\beta H2 + 1.94Vx - 1.57$	26296
Barratt M. D. (1995)	1	$\log k_p = 0.820\ \log P_{ow} - 0.0093MV - 2.36 - 0.0039Mpt$	29150
Baba H. and Ueno Y. and Hashida M. and Yamashita F. (2017)	3	$\log k_p = 4.41 \times 10\text{-}1\ LogD - 3.58 \times 10\text{-}3\ MW - 5.91$	26371
Buchwald P. and Bodor N. (2001)	1	$\log k_p = 0.0208\log Ve - 0.723N - 2.69$	29160
Cleek R. L. and Bunge A. L. (1993)	1	$k_p = klip/(1+ (klip \times \sqrt{(MW)})/2.6)$ $\log klip = -2.32 + 0.574\ \log P_{ow} - 0.005MW$	26543
Cronin M. T. and Dearden J. C. and Moss G. P. and Murray-Dickson G. (1999)	1	$\log k_p = 0.772I\ \log P_{ow} - 0.0103MW - 2.33$	26553
Cronin M. T. and Dearden J. C. and Moss G. P. and Murray-Dickson G. (1999)	2	$\log k_p = -0.0705HALP + 0.494\ \log P_{ow} - 0.230\ 4\chi v - 2.77$	26553
Dearden J. C. and Cronin M. T. D. and Patel H. and Raevsky O. A. (2000)	1	$\log k_p = -0.626\Sigma Ca - 23.8\Sigma(Q+)/a - 0.289\ SsssCH - 0.0357SsssOH - 0.482IB + 0.405BR + 0.834$	29167
Fujiwara S. I. and Yamashita F. and Hashida M. (2003)	1	$\log k_p = 0.5356\ \log P_{ow} - 0.005227MW - 2.56$	26735
Guy R. H. and Potts R. O. (1992)	1	$\log k_p = 0.74\ \log P_{ow} - 0.006MW - 2.8$	28277
Magnusson B. M. and Anissimov Y. G. and Cross S. E. and Roberts M. S. (2004)	1	$\log J_{max} = -0.0141MW - 4.52$	28496
McKone T. E. and Howd R. A. (1992)	1	$k_p = MW^{-0.6} \times [0.33 + \delta_{skin}/(-2.4\cdot10^{-6} + 3\cdot10^{-5}\ P_{OW}^{0.8})]$	29201
Mitragotri S. (2002)	1	$\log k_p = -0.20\ r^2 + 0.7\ \log P_{ow} - 1.70$	27272

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Moody R. P. and MacPherson H. (2003)	1	$\log k_p = 0.81 \log P_{ow} - 0.009$ LMV $- 2.33$	27290
Moody R. P. and MacPherson H. (2003)	2	$\log k_p = 0.77 \log P_{ow} - 0.018$ MV $- 2.25$	27290
Moody R. P. and MacPherson H. (2003)	3	$\log k_p = 0.77 \log P_{ow} - 0.010$ MW $- 2.28$	27290
Moody R. P. and MacPherson H. (2003)	4	$\log k_p = 0.81 \log P_{ow} - 0.004$ LMV $- 0.23$ STW $+ 2.06$	27290
Moss G. P. and Cronin M. T. D. (2002)	1	$\log k_p = 0.74$l $\log P_{ow} - 0.0091$MW $- 2.39$	27297
Patel H. and Berge W. T. and Cronin M. T. D. (2002)	1	$\log k_p = -2.3 + 0.652 \log P_{ow} - 0.00603$MW $-$ 0.623ABSQon $-$ 0.313SsssCH	27415
Potts R. O. and Guy R. H. (1992)	1	$\log k_p = 0.71 \log P_{ow} - 0.0061$MW $- 2.7$	28668
Potts R. O. and Guy R. H. (1995)	1	$\log k_p = 0.0256$MV $- 1.72$HD $- 3.93$HA $- 1.29$	27448
Pugh W. J. and Degim I. T. and Hadgraft J. (2000)	1	$\log k_p = -2.724 - 0.00264$ MW $\times$ charge $+ 0.59 \log P_{ow}$	27458
ten Berge W. (2009)	1	log kpSC-intercellular $= 0.7318 \log P_{ow} - 0.006832$MW $- 2.59$ log kpSC-transcellular $= -1.361$log MW $- 1.367$ k_p = kpSC-intercellular + kpSC-transcellular	27716
Ashrafi, P (2016)	1	n/a : Gaussian Process model with Matérn covariance function v=3/2	29272
Ashrafi, P (2016)	2	n/a: Support Vector Machine	29272
Atobe T. and Mori M. and Yamashita F. and Hashida M. and Kouzuki H. (2015)	1	see mixture models	26359
Atobe T. and Mori M. and Yamashita F. and Hashida M. and Kouzuki H. (2015)	2	see mixture models	26359

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Baert B. and Deconinck E. and Van Gele M. and Slodicka M. and Stoppie P. and Bode S. and Slegers G. and Vander Heyden Y. and Lambert J. and Beetens J. and De Spiegeleer B. (2007)	1	log k_p = 0.41 ALOGP + 0.61 Mor13v + 0.51 Jhetv − 1.40 Mor26v + 0.97 x P2v − 0.70 Mor11m − 0.48 MATS2e + 0.081 Mor09u − 0.11 GATS4e − 4.5	26375
Baert B. and Deconinck E. and Van Gele M. and Slodicka M. and Stoppie P. and Bode S. and Slegers G. and Vander Heyden Y. and Lambert J. and Beetens J. and De Spiegeleer B. (2007)	2	log k_p = -3.1 H.050 -1.0 Hpypertens.50 + 0.10 ALOGP - 0.00048 SRW09 + 0.15 RDF075m - 0.14 H.052 - 0.48 T.(S..F) +0.48 C.025 - 10.6 R1m+ - 6.2 RTm+ - 2.7	26375
Beydon D. and Payan J. P. and Ferrari E. and Grandclaude M. C. (2014)	1	log J_{max}, rat = –0.31 logP_{ow} + 3.7 Tlag, rat = 0.22 x exp(0.34 logP_{ow})	26426
Beydon D. and Payan J. P. and Ferrari E. and Grandclaude M. C. (2014)	2	log J_{max}, human = –0.50 logP_{ow} + 3.75 Tlag, human = 0.25 x exp(0.44 logP_{ow})	26426
Bunge A. L. and Cleek R. L. (1995)	1	k_p = klip /(1+ (klip × √(MW))/2.6) log klip = − 2.8+0.74 logP_{ow} – 0.006 MW	28068
Bunge A. L. and Flynn G. L. and Guy R. H. (1994)	1	log K(SC/W) = 0.74 logP_{ow} log (D(SC)/L(SC) = - 0.006 MW - 2.80 L(SC) = assumed to be 10-20 µm (to expressed in cm in the formula above!) For t(expo) ≤ t* : DA = 2 A C0(v) K(SC/v) √(D(SC)t(expo)/π) For t(expo) > t* : DA = A C0(v) K(SC/v) L(SC)/(1 + B) [D(SC) t(expo)/(L(SC))2 + (1 + 3 B (1 + B))/ (3 (1 + B)] t* (= time required to reach steady state) = 0.4 (L(SC))2 / D(SC) for B ≤ 0.60 t* = [b - √(b2 - c2)] (L(SC))2 / D(SC) for B > 0.60 b = 2 (1 + B)2 / π - c c = (1 + 3 B + 3 B2)/ 3 (1 + B) B = Kwp,SC x (MW)1/2 / 2.59 cm/h log Kwp,SC = 0,74 logP_{ow} - 0,0060 MW - 2.80	26482

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Chen L. J. and Lian G. P. and Han L. J. (2007)	1	log k_p = 0.34 R2 − 0.77 πH2 − 0.69 ΣαH2 − 2.32 ΣβH2 + 1.77 Vx − 2.77	26518
Chen L. J. and Lian G. P. and Han L. J. (2007)	2	n/a: ANN	26518
Dancik Y. and Miller M. A. and Jaworska J. and Kasting G. B. (2013)	1	n/a: *in silico* PBPK-model	26565
Degim T. and Hadgraft J. and Ilbasmis S. and Ozkan Y. (2003)	1	n/a: ANN	26580
Frasch H. F. (2002)	1	k_p = k_sc*kaq/(k_sc+kaq) kaq = 0.1151 cm/h k_sc = K_mv x D*/l* log K_mv = 0.59* logP_{ow} - 0.024 log D* = log_Dcor_Dlip/(1+EXP(-(log K_cor_lip + 0.1974 - 0.3668 log (Dcor/Dlip)/(0.2488 + -0.134 log (Dcor/Dlip)))) log (Dcor/Dlip) = -0.0087 MW l* = 0.003 x (1 - 0.9113 log_Kcor_lip + 0.9896 (log_Kcor_lip)^2 + 0.3111 (log_Kcor_lip)^3) log_Kcor_lip = -0.8075 logP_{ow} + 2.4194	26719
Fu X. C. and Ma X. W. and Liang W. Q. (2002)	1	log k_p = 3.69 MV - 2.86 Q(h) - 2.19 Q(O.N) - 0.033 E(HOMO) - 0.22 E(LUMO) - 1.20	26732
Fu X. C. and Ma X. W. and Liang W. Q. (2002)	2	n/a: ANN model	26732
Fu X. C. and Wang G. P. and Wang Y. F. and Liang W. Q. and Yu Q. S. and Chow M. S. (2004)	1	log k_p = 0.82 logP_{ow} - 6.83 MV - 2.38	26733
Fu X. C. and Wang G. P. and Wang Y. F. and Liang W. Q. and Yu Q. S. and Chow M. S. (2004)	2	log k_p = 0.67 logP_{ow} - 1.11 MV − 1.52 ΣαH2 − 1.21 ΣβH2 − 1.61	26733
Fu X. C. and Wang G. P. and Wang Y. F. and Liang W. Q. and Yu Q. S. and Chow M. S. (2004)	3	log k_p = 0.52 logP_{ow} -1.51 ΣαH2 − 1.45 ΣβH2 − 1.60	26733

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Fu X. C. and Wang G. P. and Wang Y. F. and Liang W. Q. and Yu Q. S. and Chow M. S. (2004)	4	log k_p = 11.6 MV -1.49 ΣαH2 – 3.87 ΣβH2 – 1.47	26733
Guth K. and Riviere J. E. and Brooks J. D. and Dammann M. and Fabian E. and van Ravenzwaay B. and Schafer-Korting M. and Landsiedel R. (2014)	1	see mixture models	26811
Katritzky A. R. and Dobchev D. A. and Fara D. C. and Hur E. and Tamm K. and Kurunczi L. and Karelson M. and Varnek A. and Solov ev V. P. (2006)	1	log k_p = -7.20 + 0.67* logP_{ow} - 0.24*Kier & Hall index (order 3) + 0.74*rotational entropy (300 K)/no. of atoms - 23.43*HASA-2/TMSA (Zefirov Partial Charge) + 0.20*number of O atoms + 3.5563	27004
Keshwani D. M. and Jones D. D. and Brand R. M. (2005)	1	n/a: fuzzy *in silico* model	27010
Khajeh A. and Modarress H. (2014)	1	log k_p = -0.877 EEig15r + 0.642 ALOGP - 1.554 Neoplastic-80 - 3.179	27015
Khajeh A. and Modarress H. (2014)	2	log k_p = -1.012 EEig15r + 0.556 logP_{ow} exp - 1.601 Neoplastic-80 - 3.084	27015
Kilian D. and Lemmer H. J. R. and Gerber M. and Du Preez J. L. and Du Plessis J. (2016)	1	log k_p = 0.739* logP_{ow} - 0.0089*MW - 2.36	27024
Krüse J. and Golden D. and Wilkinson S. and Williams F. and Kezic S. and Corish J. (2007)	1	log k_p = 0.74* logP_{ow} - 2.8 - 0.006*MW	27077
Lim C. W. and Fujiwara S. and Yamashita F. and Hashida M. (2002)	1	log k_p = -5.016 - 0.197*μ + 0.002059*pol + 0.395*sum(N,O) - 1.668*sum(H)	27150
Lindh M. and Karlén A. and Norinder U. (2017)	1	n/a: Machine learning model	27158
Lindh M. and Karlén A. and Norinder U. (2017)	2	n/a: Machine learning model	27158

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Lindh M. and Karlén A. and Norinder U. (2017)	3	n/a: Machine learning model	27158
Lindh M. and Karlén A. and Norinder U. (2017)	4	n/a: Machine learning model	27158
Luo W. and Medrek S. and Misra J. and Nohynek G. J. (2007)	1	log k_p = 0.5347* logP_{ow} - 0.09686*X0 - 0.141*SsssCH - 2.54333	27184
Luo W. and Medrek S. and Misra J. and Nohynek G. J. (2007)	2	log k_p = 0.4828* logP_{ow} - 0.003544*MW - 2.62741	27184
Luo W. and Nguyen H. and Telesford Q. and Fung W. (2004)	1	log k_p = -0.005530*MW + 0.6205* logP_{ow} + 0.000000303*Sw + 0.08239*logD(pH7.4) - 0.001483*MP - 2.759	28485
Milewski M. and Stinchcomb A. L. (2012)	1	log J_{max} = 4.600 - 0.219* logP_{ow} - 0.0086*MW - 0.0102*(MP - 25)	27255
Mitragotri S. (2003)	1	k_p = 0.0000056* P_{ow} ^0.7*exp(-0.46*r^2)	27273
Neely B. J. and Madihally S. V. and Robinson Jr R. L. and Gasem K. A. M. (2009)	1	n/a: Neural Network model	27337
Neely B. J. and Madihally S. V. and Robinson Jr R. L. and Gasem K. A. M. (2009)	2	n/a: Neural Network model	27337
Neumann D. and Kohlbacher O. and Merkwirth C. and Lengauer T. (2006)	1	n/a: machine learning model	28600
Poulin P. and Krishnan K. (2001)	1	k_p = (Pl:w*Dl*Fl)/Ll + (Pp:w*Dp*(Fp+Fw)/Lt	27449
Pugh W. J. and Hadgraft J. (1994)	1	log k_p = -0.262 + 0.335*Halide + 0.187*[C] - 0.297*[O] - 0.639*[N] + 0.275*c	29212
Pugh W. J. and Hadgraft J. (1994)	2	to be specified if model is selected as useful	29212
Pugh W. J. and Hadgraft J. (1994)	3	to be specified if model is selected as useful	29212

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Rocco P. and Cilurzo F. and Minghetti P. and Vistoli G. and Pedretti A. (2017)	1	log k_p = -1.43 + 0.72*log(P_{ow} *D/MV)	27524
Rocco P. and Cilurzo F. and Minghetti P. and Vistoli G. and Pedretti A. (2017)	2	log k_p = 0.21 + 0.75*log(P_{ow} *D300/MV^2) + 0.066*MV*(T-300)	27524
Rocco P. and Cilurzo F. and Minghetti P. and Vistoli G. and Pedretti A. (2017)	3	log k_p = -21.21 + 0.75*log(P_{ow} *D300/MV^2) + 0.072*T	27524
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	1	log k_p = 497 − 599 Jurs-FPSA-1 + 0.52 Area − 1.84 Jurs-PNSA-1	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	2	log k_p = 31.34 − 1909 Jurs-FPSA-3 +4.80 Jurs-PPSA-3 + 479 Jurs-FNSA-1 − 1.57 Jurs-PNSA-1	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	3	log k_p = 35.27 − 1.56 Jurs-PNSA-1 − 1.72 CHI-V-3_P − 2037 Jurs-FPSA-3 + 5.19 Jurs-PPSA-3 + 475 Jurs-FNSA-1	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	4	log k_p = 121 − 217 Jurs-RNCG − 37.50 JX − 7.17 CHI-V-3_P + 4.37 Jurs-RPCS + 110 Jurs-FNSA-1 + 4.02 Kappa6	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	5	log k_p = −4.85 − 0.10 (Jurs-RNCS - 5.35)2 − 688 (Jurs-RNCG - 0.21)2 + 1.34 (Jurs-RPCS - 2.91)2	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	6	log k_p = −3.20 − 0.002 (Jurs-DPSA-1 - 277)2 − 0.002 (Jurs-PNSA-1 - 40.1)2 + 0.001 (Jurs-PPSA-1 - 352)2 + 999 (Jurs-FNSA-1 - 0.11)2	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	7	log k_p = −0.49 − 0.69 (Jurs-RPCS - 3.00)2 + 0.62 (Jurs-RPCS - 2.47)2 + 702 (Jurs-RNCG - 0.21)2 − 24.9 Jurs-RPCG + 0.12 (Jurs-RNCS - 5.10)2	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	8	log k_p = −3.26 − 0.77 (Jurs-RPCS - 3.47)2 + 0.10 (Jurs-RNCS - 3.20)2 − 747 (Jurs-RNCG - 0.23)2 − 0.62 (Jurs-RPCS - 2.32)2 + 0.94 Shadow-Zlength − 38.9 Jurs-RPCG	27569
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	18	log k_p = 0.28* logP_{ow} - 0.007*MW - 2.00	27569

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Santos-Filho O. A. and Hopfinger A. J. and Zheng T. (2004)	19	$\log k_p = 0.22 * \log P_{ow} - 0.14 * Ess(tor) - 0.05 * Einter(vdW) - 2.97$	27569
Steinmetz F. P. and Madden J. C. and Cronin M. T. (2015)	1	$\log k_p = -2.51 + 0.50 * \log P_{ow} - 0.0051 * MW$	27667
Thomas J. and Majumdar S. and Wasdo S. and Majumdar A. and Sloan K. B. (2007)	1	$\log J_{max} = -2.500 + 0.5571 * logSoct + 0.4429 * logSaq - 0.00502 * MW$	27725
Thomas J. and Majumdar S. and Wasdo S. and Majumdar A. and Sloan K. B. (2007)	2	$\log J_{max} = -2.574 + 0.5861 * logSoct + 0.4139 * logSaq - 0.00440 * MW$	27725
Walker J. D. and Rodford R. and Patlewicz G. (2003)	1	$\log k_p = 0.595 * \log P_{ow} - 3.212$	27803
Williams F. M. and Rothe H. and Barrett G. and Chiodini A. and Whyte J. and Cronin M. T. and Monteiro-Riviere N. A. and Plautz J. and Roper C. and Westerhout J. and Yang C. and Guy R. H. (2016)	1	$J_{max} = Kp,mod * Caq,sat$	27847
Williams F. M. and Rothe H. and Barrett G. and Chiodini A. and Whyte J. and Cronin M. T. and Monteiro-Riviere N. A. and Plautz J. and Roper C. and Westerhout J. and Yang C. and Guy R. H. (2016)	2	$Qmax = A * J_{max} * Texp$	27847
Xu G. and Hughes-Oliver J. M. and Brooks J. D. and Yeatts J. L. and Baynes R. E. (2013)	1	$\log k_p = -2.50 + 0.04 * Ei + 0.83 * Si - 0.13 * Ai - 1.00 * Bi + 0.27 * Vi$	27872
Yu Y. J. and Su R. X. and Wang L. B. and Qi W. and He Z. M. (2010)	1	$\log k_p = -1.56 - 0.153 * Dipole - 0.259 * alog P_{ow} + 0.0385 * Har - 0.543 * Xu + 0.976 * \log P_{ow} - 0.00626 * Mp$	27901

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Yu Y. J. and Su R. X. and Wang L. B. and Qi W. and He Z. M. (2010)	2	$\log k_p$ = -0.255 - 0.127*Dipole + 24.6*CP + 0.053*MolP + 0.014*Har - 1.94*AE - 1.29*Xu + 0.080*LSI - 0.00186*TE - 0.00413*Mv + 0.774* $\log P_{ow}$ - 0.00691*Mp	27901
Zhang K. and Abraham M. H. and Liu X. (2017)	1	$\log k_p$ = -5.328 + 0.137*E - 0.604*S - 0.338*A - 2.428*B + 1.797*V - 1.485*J+ + 2.471*J-	29242
Zhang K. and Chen M. and Scriba G. K. E. and Abraham M. H. and Fahr A. and Liu X. (2012)	1	$\log k_p$ = -5.420 - 0.102*E - 0.457*S - 0.324*A - 2.680*B + 2.066*V - 1.938*J+ + 2.548*J-	27918

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table D.2: Algorithm models possibly relevant for pesticide risk assessment - Mixture models.

Citation	Model No.	Model algorithm(s)	LRT-ID
Chittenden J. T. and Riviere J. E. (2016)	1	$\log k_p$ = -7.245 - 0.6775 ($\log P_{ow}$ - 2.910) + 0.00832 (MW-225) - 0.0304 (mCMA - 75.3) - 0.0229 (mTPSAd - 27.3) + 0.778	26534
(Ghafourian *et al.*, 2010b)	1	$\log k_p$ = −0.956 − 0.00322 Δmp − 0.000320W(P) − 0.0121 BP(V) − 0.114 Lipole(P)	26771
Ghafourian T. and Samaras E. G. and Brooks J. D. and Riviere J. E. (2010)	2	$\log k_p$ = −310 − 0.000315W(P) − 0.00771 δ(V) x E(HOMO)(P) − 0.0102 BP(V) − 0.0750 Lipole(P)	26771
Ghafourian T. and Samaras E. G. and Brooks J. D. and Riviere J. E. (2010)	3	$\log k_p$ = −2.48 − 0.0474 N(atoms)(P) − 0.00798 δ(V) x E(HOMO)(P) − 0.0102 BP(V) − 0.0723 Lipole(P)	26771
Ghafourian T. and Samaras E. G. and Brooks J. D. and Riviere J. E. (2010)	4	$\log k_p$ = −4.29 − 0.0474 N(atoms)(P) − 0.00904 (BP − MP)(V) − 0.345 E(HOMO)(P) − 0.0790 Lipole(P)	26771
Riviere J. E. and Brooks J. D. (2005)	1	$\log k_p$ = - 10.394 mRI − 1.527 ΣαH2 + 0.045 ΣβH2 + 0.327 πH2 - 0.561 R2 - 1.904 Vx + 13.921	27505
Riviere J. E. and Brooks J. D. (2005)	2	$\log k_p$ = - 9.242 mRI − 0.525 ΣαH2 + 0.329 ΣβH2 + 0.407 πH2 - 0.411 R2 - 1.385 Vx + 10.751	27505
Riviere J. E. and Brooks J. D. (2005)	3	$\log k_p$ = - 0.318 mPo − 1.529 ΣαH2 + 0.043 ΣβH2 + 0.327 πH2 - 0.563 R2 - 1.902 Vx + 0.81	27505
Riviere J. E. and Brooks J. D. (2005)	4	$\log k_p$ = - 0.149 mPo − 0.155 ΣαH2 + 0.546 ΣβH2 + 0.421 πH2 - 0.433 R2 - 1.255 Vx - 1.255	27505
Riviere J. E. and Brooks J. D. (2005)	5	$\log k_p$ = - 0.479 mlog(1/HC) − 1.282 ΣαH2 + 0.195 ΣβH2 + 0.280 πH2 - 0.453 R2 - 1.863 Vx + 2.372	27505
Riviere J. E. and Brooks J. D. (2005)	6	$\log k_p$ = - 0.419 mlog(1/HC) − 0.050 ΣαH2 + 0.693 ΣβH2 + 0.388 πH2 - 0.323 R2 - 1.328 Vx + 0.105	27505
Riviere J. E. and Brooks J. D. (2007)	1	$\log k_p$ = -0.04 mTPSA - 0.03 $\log P_{ow}$ - 0.00080 MW - 2.05	27506
Riviere J. E. and Brooks J. D. (2007)	2	$\log k_p$ = -1.19 mHA - 0.03 $\log P_{ow}$ - 0.00081 MW - 1.12	27506
Riviere J. E. and Brooks J. D. (2007)	3	$\log k_p$ = -0.04 mTPSA - 0.48 MR + 0.09 HBA - 0.42 HBD - 0.49	27506
Riviere J. E. and Brooks J. D. (2007)	4	$\log k_p$ = -10.42 mRI - 0.49 MR + 0.10 HBA - 0.45 HBD + 13.01	27506
Riviere J. E. and Brooks J. D. (2007)	5	$\log k_p$ = - 0.04 mTPSA − 1.45 ΣαH2 + 0.31 ΣβH2 + 0.0.004 πH2 - 0.45 R2 - 1.92 Vx + 0.54	27506

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
Riviere J. E. and Brooks J. D. (2007)	6	log k_p = - 10.38 mRI − 1.54 ΣαH2 + 0.25 ΣβH2 + 0.092 πH2 - 0.53 R2 - 1.92 Vx + 13.98	27506
Riviere J. E. and Brooks J. D. (2011)	1	log k_p = − 1.45 ΣαH2 + 0.01 ΣβH2 + 0.27 πH2 - 0.55 R2 - 1.39 Vx + 2.55	27508
Riviere J. E. and Brooks J. D. (2011)	2	log k_p = 77.66/mMP − 1.47 ΣαH2 + 0.01 ΣβH2 + 0.27 πH2 - 0.56 R2 - 1.39 Vx + 3.14	27508
Riviere J. E. and Brooks J. D. (2011)	3	log k_p = -1.21 mHbacc − 1.40 ΣαH2 + 0.03 ΣβH2 + 0.27 πH2 - 0.55 R2 - 1.38 Vx + 4.25	27508
Riviere J. E. and Brooks J. D. (2011)	4	log k_p = -1.21 mHbacc − 1.44 ΣαH2 + 0.28 πH2 - 0.55 R2 - 1.38 Vx + 4.27	27508
Riviere J. E. and Brooks J. D. (2011)	5	log k_p = −0.10 logP_{ow} − 0.00058 MW + 1.13	27508
Riviere J. E. and Brooks J. D. (2011)	6	log k_p = −1.23 mHBacc − 0.10 logP_{ow} − 0.00058 MW + 2.89	27508
Riviere J. E. and Brooks J. D. and Collard W. T. and Deng J. and de Rose G. and Mahabir S. P. and Merritt D. A. and Marchiondo A. A. (2014)	1	log k_p = - 1.43 mlogP_{ow} − 14.86 ΣαH2 + 2.02 ΣβH2 + 2.97 πH2 - 11.86 R2 + 8.87	27509
Riviere J. E. and Brooks J. D. and Collard W. T. and Deng J. and de Rose G. and Mahabir S. P. and Merritt D. A. and Marchiondo A. A. (2014)	2	log k_p = - 1.39 mlog WS(M) − 14.17 ΣαH2 + 1.64 ΣβH2 + 2.93 πH2 - 11.80 R2 + 9.30	27509
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	1	log J_{SS} = 0.00235 m(BP - MP) + 0.000001 [donor] - 0.00570 MW + 3.96 vsurf_G + 0.0137 S log P_VSA4 - 1.93 fiAB - 0.343 VAdjMa - 1.92	27563
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	2	log J_{SS} = 0.00192 m(BP - MP) + 0.000001 [donor] - 0.00561 MW + 3.82 vsurf_G + 0.0140 S log P_VSA4 - 1.95 fiAB - 0.312 VAdjMa - 1.67 - 0.201 Thickness	27563
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	3	log J_{SS} = 0.00230 m(BP - MP) + 0.000001 [donor] - 0.00592 MW + 3.55 vsurf_G + 0.00992 S log P_VSA4 - 1.85 fiAB - 0.293 VAdjMa - 1.08 - 0.391 Infinite/Finite	27563
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	4	n/a: Regression Tree model based on parameters listed in previous column	27563
Samaras E. G. and Riviere J.	5	n/a: Regression Tree model based on parameters listed in previous column	27563

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Citation	Model No.	Model algorithm(s)	LRT-ID
E. and Ghafourian T. (2012)			
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	6	n/a: Regression Tree model based on parameters listed in previous column	27563
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	7	n/a: Regression Tree model based on parameters listed in previous column	27563
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	8	n/a: Regression Tree model based on parameters listed in previous column	27563
Samaras E. G. and Riviere J. E. and Ghafourian T. (2012)	9	n/a: Regression Tree model based on parameters listed in previous column	27563
Atobe T. and Mori M. and Yamashita F. and Hashida M. and Kouzuki H. (2015)	1	$\log k_p = -0.193 \times \log P_{ow}$ (chemical) $\times \log P_{ow}$ (vehicle) + 0.00124 x MW(chemical) $\times \log P_{ow}$ (vehicle) - 0.00476 x MW(chemical) + 0.0184 x $(\log P_{ow}$ (vehicle))^2 - 0.00000352 x MW(chemical)^2 - 2.23	26359
Atobe T. and Mori M. and Yamashita F. and Hashida M. and Kouzuki H. (2015)	2	n/a: ANN using MW(chemical), P_{ow} (chemical), P_{ow} (vehicle)	26359
Guth K. and Riviere J. E. and Brooks J. D. and Dammann M. and Fabian E. and van Ravenzwaay B. and Schafer-Korting M. and Landsiedel R. (2014)	1	log (max k_p) = 0.6 Rf2 + 0.5 SpI - 0.1 mTPSA- 2.3	26811
(Ghafourian *et al.*, 2010a)	1	$\log k_p = -0.909 - 0.610 \log P_{ow}$+ 2.62 (^9)χ(p) −0.00917 (SolBP − SolMP)	26770

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table D.3: Parameter codes.

Code	Description
A	Effective/overall hydrogen bond acidity, Abraham descriptor
α	hydrogen bond donor acidity
ABSQon	the sum of absolute charges on oxygen and nitrogen atoms
accptHB	QikProp descriptor
AE	average of estate values
ΣαH2	Effective/overall hydrogen bond acidity, Abraham descriptor
ALOGP	Ghose-Crippen octanol-water partition coeff. (logP) (from Dragon)
aLogP	hydrophobicity, probably the same as ALOGP
Area	molecular surface area: a 3D spatial descriptor that describes the van der Waals area of a molecule
B	Effective/overall hydrogen bond basicity, Abraham descriptor
B03[O-O]	Presence/absence of O - O at topological distance 3 (from Dragon)
B04[C-O]	Presence/absence of C - O at topological distance 4 (from Dragon)
B05[O-O]	Presence/absence of O - O at topological distance 5 (from Dragon)
B06[C-N]	Presence/absence of C - N at topological distance 6 (from Dragon)
ΣβH2	Effective/overall hydrogen bond basicity, Abraham descriptor
BLTF96	Verhaar Fish base-line toxicity from MLOGP (mmol/l) (from Dragon)
Bp	=BP = boiling point
BP - MP	BP – MP is the difference between the boiling and melting points of a compound
BR	Number of rotable bonds
c	aromatic carbon atom
[C]	single-bonded carbon atom
C.025	the atom-centred fragment R-CR-R
C0(v)	concentration of penetrant in vehicle (mg/cm3)
CATS2D_06_LL	CATS2D Lipophilic-Lipophilic at lag 06 (from Dragon)
cb	the number of carbons not involved in a C= O bond
charge	the sum of the absolute values of the partial charges (calculation method described in Pugh *et al.* (2000))
chi0	Zero order molecular connectivity index (Hall and Kier (1991))
chi0V	Zero order valence molecular connectivity index (Hall and Kier (1991))
chi1v_C	First order carbon valence connectivity index (Hall and Kier (1991))
(^9)χ(p)	9th order path molecular connectivity index of the penetran
ChiA_H2	average Randic-like index from reciprocal squared distance matrix (from Dragon)
4χv	fourth-order valence-corrected molecular connectivity
CHI-V-3_P	Kier and Hall valence-modified connectivity index CHI-3_P means third-order CHI index with three paths (bonds) connected; V means that electron configuration of the atom (single or multiple bonds) is considered

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
Count of H-acceptor sites [Zefirov] (CODESSA)	calculated by CODESSA
Count of H-donor sites [Zefirov]	calculated by CODESSA
CP	charge polarization
Csat	Concentration in water at saturation (expressed in mg/cm3)
δ	solubility parameter, calculated based on Fedor's approach (see: K.C. James, in: J. Swarbrick (Ed.), Solubility and Related Properties, Vol. 5, Marcel Dekker, New York, 1986, pp. 184–188)
δ	the Hildebrand solubility parameter
ddHsolv	solvation enthalpy
den_2	not specified, probably density in µg/mL
Dipole	dipole moment
Dl	lipid diffusion coefficient
DLS_05	modified drug-like score from Zheng *et al.* (2 rules) (from Dragon)
Δmp	the difference between the melting point of the penetrant and that of the solvent
[donor]	donor concentration a.i. (µg/mL)
donorHB	QikProp descriptor
Dp	protein diffusion coefficient
δskin	skin thickness in cm (0.0025 cm in McKone and Howd (1992))
E	Excess molar refraction, Abraham descriptor
E(HOMO)	the energy of the highest occupied molecular orbital (eV)
E(LUMO)	the energy of the lowest unoccupied molecular orbital (eV)
E1v	WHIM descriptor, calculated by DRAGON
E2s	2nd component accessibility directional WHIM index / weighted by I-state (from Dragon)
EA(eV)	QikProp descriptor
EEig15r	an edge adjacency index as a topological descriptor derived from the eigenvalue of the adjacency matrix of edges weighted by resonance integrals [33]
Eig04_AEA(ed)	eigenvalue n. 4 from augmented edge adjacency mat. weighted by edge degree (from Dragon)
Eig10 EA(ed)	eigenvalue n. 10 from edge adjacency mat. weighted by edge degree (from Dragon)
Eta beta A	eta average VEM count (from Dragon)
fiAB	Fraction of molecules ionised as anion and cation at pH 7.4
Fl	lipid fractional content
Fp	protein fractional content
Fw	water fractional content
G2i	2nd component symmetry directional WHIM index / weighted by ionization potential (from Dragon)
GATS4e	three-dimensional (Geary autocorrelation lag 4)spatial autocorrelation coefficient
GATS5e	Geary autocorrelation of lag 5 weighted by Sanderson electronegativity (from Dragon)
GCUT_PEOE_1	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
	the adjacency matrix takes the value 1/sqr(dij) where dij is the (modified) graph distance between atoms i and j. The diagonal takes the value of the PEOE partial charges. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and the largest eigenvalues are reported (MOE (2011))
GCUT_SLOGP_1	The GCUT descriptors using atomic contribution to $logP_{ow}$ instead of partial charge (MOE (2011))
GCUT_SMR_0	The GCUT descriptors using atomic contribution to molar refractivity using the instead of partial charge (MOE (2011))
GCUT_SMR_3	The GCUT descriptors using atomic contribution to molar refractivity instead of partial charge (MOE (2011))
glob	QikProp descriptor
H.050	the number of hydrogen atoms attached to a heteroatom (atom-centred fragment)
H.052	the number of hydrogen's attached to C(sp3) with 1 halogen attached to the next C
H3p	H autocorrelation of lag 3 / weighted by polarizability (from Dragon)
HA	Hydrogen bond acceptor activity
HALP	total number of lone pairs that can accept hydrogen bonds
Har	Harary index
HASA-2/TMSA	Zefirov Partial Charge
Hb	number of hydrogens
HBA	hydrogen bond acceptors = counts of hydrogen bond donors
HBD	hydrogen bond donors = counts of hydrogen bond donors
HD	Hydrogen bond donor activity
HeavyAtomCount	number of non-hydrogen atoms
Hypertens.50	the Ghose–Viswanadhan–Wendoloski 50%-antihypertensive drug-like index (molecular property class)
IB	Balaban index
Infinite/Finite	Indicator variable indicating infinite or finite exposures taking a value of 2 for finite and 1 for infinite dosing
Jhetv	Balaban type index obtained from the van der Waals volume weighed distance matrix
Jurs-DPSA-1	difference in charged partial surface areas: partial positive solvent-accessible surface area minus partial negative solvent-accessible surface area
Jurs-FNSA-1	fractional charged partial surface areas: total charge weighted negative surface area divided by the total molecular solvent-accessible surface area
Jurs-FPSA-1	fractional charged partial surface areas: partial positive surface area divided by the total molecular solvent-accessible surface area
Jurs-FPSA-3	fractional charged partial surface areas: total charge weighted positive surface area divided by the total molecular solvent-accessible surface area
Jurs-PNSA-1	partial negative surface area: sum of the solvent-accessible surface areas of all negatively charged atoms
Jurs-PPSA-1	partial positive surface area: sum of the solvent-accessible surface areas of all positively charged atoms
Jurs-PPSA-3	atomic charge weighted positive surface area: Sum Of the product Of solvent-accessible surface area X partial charge for all posltlvely charged atoms
Jurs-RNCG	relative negative charge: charge of most negative atom divided by the total negative charge

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
Jurs-RNCS	relative negative charge surface area: solvent-accessible surface area of most negative atom divided by relative negative charge
Jurs-RPCG	relative positive charge: charge of most positive atom divided by the total positive charge
Jurs-RPCS	relative positive charge surface area: solvent accessible surface area of most positive atom divided by descriptor
JX	Balaban index: characterizes the shape of a molecule, which can take account of the covalent radii
Kappa6	Kiers shape indices; Kappa6 is the sixth-order index, compares the molecule graph with "minimal" and "maximal" graphs
Kier shape index – order 1	calculated by CODESSA
KierA1	First alpha modified shape index, also correlated with molecular size (Hall and Kier (1991))
KierA3	Third alpha modified shape index, informing centrality of branching with large values representing location of branching at the extremities of the molecule (Hall and Kier (1991))
P_{ow}	octanol-water partition coefficient
L(SC)	thickness of *stratum corneum* in cm
L3s	WHIM descriptor, calculated by DRAGON
Linearity	not further specified
Lipole	the total lipole moment of the penetrants
LI	intercellular lipid pathway
LMV	Liquid Molar Volume
$logP_{ow}exp$	experimentally determined $LogP_{ow}$
logD	pH dependent $LogP_{ow}$
logS	Log of the aqueous solubility (mol/l) calculated by MOE from an atom contribution linear atom type model (MOE (2011))
LSI	log of superpendentic index
Lt	transcellular proteinaceous pathway
LUMO energy	lowest unoccupied molecular orbital energy calculated by CODESSA
μ	dipole moment
m log $WS_{(M)}$	mixture log Water Solibility (expressed in Mol/L), WS obtained via http://www.vcclab.org/lab/alogps/
m(BP-MP)	mixture: weighted difference between the boiling and melting points of its compounds (excluidng a.i.'s)
MATS 1m	Moran autocorrelation of lag 1 weighted by mass (from Dragon)
MATS 1p	Moran autocorrelation of lag 1 weighted by polarizability (from Dragon)
MATS2e	two-dimensional (Moran autocorrelation lag 2) spatial autocorrelation coefficient
$maxk_p$	maximum Kp
mCMA	= mixture Conolly molecular area = concentration weighted average (based on contribution to total formula weight) of Conolly moleculare areas of mixture constituent molecules
mHA	mixture number of hydrogen bond acceptors (HA calulated by log *P*.com)
mHBacc	mixture number of hydrogen bond acceptors (HA calulated by www.molinspiration.com)
$mLogP_{ow}$	mixture $LogP_{ow}$
mlog WS(M)	mixture log water solubility (expressed as moles/L)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
mlog(1/HC)	mixture log(1/Henry's Law Constant)
mMP	mixture melting point
MOE CLOGP	MOE property $LogP_{ow}$, calculated with Molecular Operating Environment program
MolP	molecular polarizability
Mor09u	descriptor of 3D-Molecule Representation of Structures based on Electron diffraction, unweighed
Mor11m	descriptor of 3D-Molecule Representation of Structures based on Electron diffraction, weighed by atomic masses
Mor13v	descriptor of 3D-Molecule Representation of Structures based on Electron diffraction, weighed by van der Waals volumes
Mor26v	descriptor of 3D-Molecule Representation of Structures based on Electron diffraction, weighed by van der Waals volumes
Mor28i	signal 28 / weighted by ionization potential (from Dragon)
Mor32s	signal 32 / weighted by I-state (from Dragon)
MP	melting point
MP (°C)	melting point in degree Celsius
mPo	mixture Polarizability
Mpt	melting point
MR	Molar Refractivity
MR	molecular refractivity
mRI	mixture Refractive Index (RI calculated by SPARC)
mTPSA	mixture Topological Polar Surface Area = concentration weighted average (based on contribution to total formula weight) of polar surface areas of mixture constituent molecules: m = weight fraction of ingredient in formulation p = property of ingredient in formulation (in this case TPSA, calculated by ADME) $MF = \sum_i m_i p_i$
mTPSAd	mixture Topological Polar Surface Area difference = TPSA a.i. minus concentration weighted average (based on contribution to total formula weight) of polar surface areas of mixture constituent molecules: MFd = mixture factor difference p(j) = a.i. $MFd = p_j - \sum_i m_i p_i$
Mv	molecular volume
MV	molecular volume (nm^3)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
MV	molar volume
MW	molecular weight
MW	molecular weight
[N]	any nitrogen atom
N	Number of affected hydrogen bonds, calculated by summing the number of N and O atoms (aliphatic twice, aromatic once) (Bodor and Buchwald 1997)
N(atoms)	the total number of atoms in the molecules
nArCOOR	number of esters (aromatic), calculated by DRAGON
nCq	number of total quaternary C(sp3), calculated by DRAGON
Neoplastic-80	antineoplastic-like property filter at 80% similarity, a descriptor proposed by Ghose *et al.* [34]. It is a set of general and objective rules based on the limits of structural features and physicochemical properties.
nHAcc	number of acceptor atoms for H-bonds (N,O,F) (from Dragon)
nHDon	number of hydrogen donors, calculated by DRAGON
nRCOOR	number of esters (aliphatic) (from Dragon)
Number of single bonds	calculated by CODESSA
NumRotatableBonds	number of rotatable bonds
[O]	any oxygen atom
Occlusion	Indicator variable for occlusion of the skin during *in vitro* test
P2v	one of the directional WHIM (Weighted Holistic Invariant Molecular) descriptors, capturing molecular 3D information related to the shape of the molecule. P2v represents the 2nd component shape directional WHIM index, weighed by the atomic van der Waals volumes
PEOE_RPC+	Relative positive partial charge: the largest positive atomic partial charge divided by the sum of the positive partial charges (MOE 2011))
PEOE_VSA_POL	Could be a printing error as the term is not in the list of seection 5 of Samaras *et al.* (2012) but PEOE_VSA_POS is.
PEOE_VSA_POS	Total positive van der Waals surface area. This is the sum of the van der Waals surface area of atoms with non-negative partial charges (MOE 2011))
	Solute dipolarity/polarizability, Abraham descriptor
PISA	QikProp descriptor
Pl:w	lipid:water partition coefficient
pol	polarizability
Polarity parameter (Qmax-Qmin)	calculated by CODESSA
Pp:w	protein:water partition coefficient
pre-Hydration	Indicator variable for pre-hydration of the skin prior to the *in vitro* test
PSA	polar surface area
Q(h)	the sum of the net atomic charges of the hydrogen atoms bound to nitrogn or oxygen atoms
Q(O.N)	the sum of the absolute values of the net atomic charges of oxygene and nitrogen atoms which were hydrogen-bond

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
	acceptors
r	van der Waal's molecular radius in Å(ngstrom) (10-10 m)
R1m	R autocorrelation of lag 1 / weighted by mass (from Dragon)
R1m+	R maximal autocorrelation of lag 1, weighted by atomic masses
R2	Excess molar refraction, Abraham descriptor
R2	Excess molar refraction, Abraham descriptor
R4p	R autocorrelation of lag 4 / weighted by polarizability (from Dragon)
R7s	R autocorrelation of lag 7 / weighted by I-state (from Dragon)
RDF070s	Radial Distribution Function - 070 / weighted by I-state (from Dragon)
RDF075m	the radial distribution function 7.5 weighted by atomic masses
RDF090i	Radial Distribution Function - 090 / weighted by ionization potential (from Dragon)
RDF1 10i	Radial Distribution Function -110/ weighted by ionization potential (from Dragon)
RDF120s	Radial Distribution Function - 120 / weighted by I-state (from Dragon)
Rf2	excess molar refraction
RingCount	number of rings
#rotor	Number of rotable bonds
RTm+	R maximal index, weighted by atomic masses
S	Solute dipolarity/polarizability, Abraham descriptor
S lo P_vSA4	sum of van der Waals surface area of atoms with $logP_{ow}$ contributions in the range of (0.1–0.15) (MOE (2011))
S log P_VSA4	sum of van der Waals surface area of atoms with log P ($=P_{ow}$) contributions in the range of (0.1–0.15) (MOE (2011))
Shadow-Zlength	length of molecule in the Z dimension
SM 1_Dz(v)	spectral moment of order 1 from Barysz matrix weighted by van der Waals volume (from Dragon)
SOLV [kJ/mol]	solvation free energy
SP	solubility parameter expressed in (cal/cm3)1/2
SpI	Species Indicator varibale (rat = 2, human = 1)
SpMaxA B(s)	normalized leading eigenvalue from Burden matrix weighted by I-State (from Dragon)
SpMaxA_EA(ri)	normalized leading eigenvalue from edge adjacency mat. weighted by resonance integral (from Dragon)
SRW09	the self-returning walk count of order 09
SsssCH	the sum of E-state indices for all methyl groups/Electrotopological atom-type index for singly bonded CH
SsssOH	Electrotopological atom-type index for singly bonded OH
STW	Surface Tension in Water
sum(H)	sum of charges of hydrogen atoms bonding to nitrogen or oxygen atoms
sum(N,O)	sum of charges of nitrogen and oxygen atoms
∑Ca	HYBOT-PLUS H-bond acceptor free energy factor
∑(Q+)/α	HYBOT-PLUS positive charge per unit volume

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
SolBP – SolMP	difference between the boiling point and the melting point of the solvent system.
Sw	water solubility
t(expo)	exposure time in hours
T.(S..F)	the sum of topological distances between S and F atoms
TDB05v	3D Topological distance based descriptors - lag 5 weighted by van der Waals volume (from Dragon)
TDB09e	3D Topological distance based descriptors - lag 9 weighted by Sanderson electronegativity (from Dragon)
TE	total energy
Thickness	Thickness of the membrane expressed in mm (millimeters)
Total dipole of the molecule	calculated by CODESSA
TPSA	Topological polar surface area
V	McGowan characteristic volume, Abraham descriptor
VAdjMa	Vertex adjacency information which depends on the number of heavy-heavy bonds (Cruciani *et al.* (2000))
Ve	Van der Waals effective molecular volumes, calculated according to Buchwald and Bodor (1998)
VE1_H2	coefficient sum of the last eigenvector from reciprocal squared distance matrix (from Dragon)
VEH	vehicle type (for acetone VEH=1 and for ethanol VEH=2)
vsa_acc	Approximation to the sum of VDW surface areas of pure hydrogen bond acceptors (not counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as OH) (MOE (2011))
vsa_hyd	Approximation to the sum of VDW surface areas of hydrophobic atoms (MOE (2011))
vsurf_CW3	Capacity factor representing the ratio of the hydrophilic surface over the total molecular surface. These are calculated at eight different energy levels. (Cruciani *et al.* (2000))
vsurf_D6	Volume that can generate hydrophobic interactions. VolSurf computes hydrophobic descriptors at eight different energy levels (Cruciani *et al.* (2000))
vsurf_EWmin1	The lowest hydrophilic interaction energy
vsurf_G	The molecular globularity–how spherical a molecule is, where values above 1 is non-perfect spheres (Cruciani *et al.* (2000))
vsurf_HB5	H-bond donor capacity, representing the molecular envelope which can generate attractive H-donor interactions with carbonyl oxygen probe. The descriptors are computed at six different energy levels (Cruciani *et al.* (2000))
vsurf_W1	Hydrophilic volume describing the molecular envelope which attractively interacts with water molecules at eight different energy levels (Cruciani *et al.* (2000))
vsurf_Wp2	Polar volume (Cruciani *et al.* (2000))
Vx	McGowan characteristic volume, Abraham descriptor
Vx	McGowan characteristic volume, Abraham descriptor
W	the Wiener topological index (the sum of distances between all pairs of vertices in the molecular graph of an alkane)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Code	Description
weinerPath	Wiener path number: half sum of all the distance matrix entries (MOE (2011))
WS	water solubility (mg/L)
X0	zero order molecular connectivity chi index
Xu	Xu index

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Full reference list of models possibly relevant for pesticide risk assessment (relates to Table D.1 & Table D.2)

Abraham, M. H. and F. Martins (2004). "Human skin permeation and partition: General linear free-energy relationship analyses." Journal of Pharmaceutical Sciences 93(6): 1508-1523.

Abraham, M. H., F. Martins and R. C. Mitchell (1997). "Algorithms for skin permeability using hydrogen bond descriptors: The problem of steroids." Journal of Pharmacy and Pharmacology 49(9): 858-865.

Ashrafi, P. (2016). Prediciting the absorption rate of chemicals through mammalian skin using machine learning algorithms. PhD, University of Hertfordshire.

Atobe, T., M. Mori, F. Yamashita, M. Hashida and H. Kouzuki (2015). "Artificial neural network analysis for predicting human percutaneous absorption taking account of vehicle properties." J Toxicol Sci 40(2): 277-294.

Baba, H., Y. Ueno, M. Hashida and F. Yamashita (2017). "Quantitative prediction of ionization effect on human skin permeability." International Journal of Pharmaceutics 522(1-2): 222-233.

Baert, B., E. Deconinck, M. Van Gele, M. Slodicka, P. Stoppie, S. Bodé, G. Slegers, Y. Vander Heyden, J. Lambert, J. Beetens and B. De Spiegeleer (2007). "Transdermal penetration behaviour of drugs: CART-clustering, QSPR and selection of model compounds." Bioorganic and Medicinal Chemistry 15(22): 6943-6955.

Barratt, M. D. (1995). "Quantitative structure-activity relationships for skin permeability." Toxicology in Vitro 9(1): 27-37.

Beydon, D., J. P. Payan, E. Ferrari and M. C. Grandclaude (2014). "Percutaneous absorption of herbicides derived from 2,4-dichlorophenoxyacid: Structure-activity relationship." Toxicology in Vitro 28(5): 1066-1074.

Buchwald, P. and N. Bodor (2001). "A simple, predictive, structure-based skin permeability model." J Pharm Pharmacol 53(8): 1087-1098.

Bunge, A. L. and R. L. Cleek (1995). "A New Method for Estimating Dermal Absorption from Chemical Exposure: 2. Effect of Molecular Weight and Octanol-Water Partitioning." Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists 12(1): 88-95.

Bunge, A. L., G. L. Flynn and R. H. Guy (1994). Predictive model for dermal exposure assessment. Wang, R. G. M. (Ed.). Environmental Science and Pollution Control Series, 9. Water Contamination and Health: Integration of Exposure Assessment, Toxicology, and Risk Assessment. Xiv+524p. Marcel Dekker, Inc.: New York, New York, USA; Basel, Switzerland. Isbn 0-8247-8922-9.; 0 (0). 1994. 347-373.

Chen, L. J., G. P. Lian and L. J. Han (2007). "Prediction of human skin permeability using artificial neural network (ANN) modeling." Acta Pharmacol Sin 28(4): 591-600.

Chittenden, J. T. and J. E. Riviere (2016). "Assessment of penetrant and vehicle mixture properties on transdermal permeability using a mixed effect pharmacokinetic model of ex vivo porcine skin." Biopharmaceutics and Drug Disposition 37(7): 387-396.

Cleek, R. L. and A. L. Bunge (1993). "A New Method for Estimating Dermal Absorption from Chemical Exposure. 1. General Approach." Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists 10(4): 497-506.

Cronin, M. T. D., J. C. Dearden, G. P. Moss and G. Murray-Dickson (1999). "Investigation of the mechanism of flux across human skin in vitro by quantitative structure-permeability relationships." European Journal of Pharmaceutical Sciences 7(4): 325-330.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Dancik, Y., M. A. Miller, J. Jaworska and G. B. Kasting (2013). "Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure." Advanced Drug Delivery Reviews 65(2): 221-236.

Dearden, J. C., M. T. D. Cronin, H. Patel and O. A. Raevsky (2000). "QSAR prediction of human skin permeability coefficients." Journal of Pharmacy and Pharmacology 52(9 SUPPL.): 221.

Degim, T., J. Hadgraft, S. Ilbasmis and Y. Ozkan (2003). "Prediction of skin penetration using artificial neural network (ANN) modeling." J Pharm Sci 92(3): 656-664.

Frasch, H. F. (2002). "A random walk model of skin permeation." Risk Anal 22(2): 265-276.

Fu, X. C., X. W. Ma and W. Q. Liang (2002). "Prediction of skin permeability using an artificial neural network." Pharmazie 57(9): 655-656.

Fu, X. C., G. P. Wang, Y. F. Wang, W. Q. Liang, Q. S. Yu and M. S. S. Chow (2004). "Limitation of Potts and Guy's model and a predictive algorithm for skin permeability including the effects of hydrogen-bond on diffusivity." Pharmazie 59(4): 282-285.

Fujiwara, S., F. Yamashita and M. Hashida (2003). "QSAR analysis of interstudy variable skin permeability based on the "latent membrane permeability" concept." J Pharm Sci 92(10): 1939-1946.

Ghafourian, T., E. G. Samaras, J. D. Brooks and J. E. Riviere (2010). "Modelling the effect of mixture components on permeation through skin." International Journal of Pharmaceutics 398(1-2): 28-32.

Ghafourian, T., E. G. Samaras, J. D. Brooks and J. E. Riviere (2010). "Validated models for predicting skin penetration from different vehicles." European Journal of Pharmaceutical Sciences 41(5): 612-616.

Guth, K., J. E. Riviere, J. D. Brooks, M. Dammann, E. Fabian, B. van Ravenzwaay, M. Schafer-Korting and R. Landsiedel (2014). "In silico models to predict dermal absorption from complex agrochemical formulations." SAR QSAR Environ Res 25(7): 565-588.

Guy, R. H. and R. O. Potts (1992). "Structure-permeability relationships in percutaneous penetration." Journal of Pharmaceutical Sciences 81(6): 603-604.

Katritzky, A. R., D. A. Dobchev, D. C. Fara, E. Hür, K. Tämm, L. Kurunczi, M. Karelson, A. Varnek and V. P. Solov'ev (2006). "Skin permeation rate as a function of chemical structure." Journal of Medicinal Chemistry 49(11): 3305-3314.

Keshwani, D. M., D. D. Jones and R. M. Brand (2005). "Takagi-Sugeno fuzzy modeling of skin permeability." Cutaneous and Ocular Toxicology 24(3): 149-163.

Khajeh, A. and H. Modarress (2014). "Linear and nonlinear quantitative structure-property relationship modelling of skin permeability." SAR QSAR Environ Res 25(1): 35-50.

Kilian, D., H. J. Lemmer, M. Gerber, J. L. du Preez and J. du Plessis (2016). "Exploratory data analysis of the dependencies between skin permeability, molecular weight and log P." Pharmazie 71(6): 311-319.

Krüse, J., D. Golden, S. Wilkinson, F. Williams, S. Kezic and J. Corish (2007). "Analysis, interpretation, and extrapolation of dermal permeation data using diffusion-based mathematical models." Journal of Pharmaceutical Sciences 96(3): 682-703.

Lim, C. W., S. I. Fujiwara, F. Yamashita and M. Hashida (2002). "Prediction of human skin permeability using a combination of molecular orbital calculations and artificial neural network." Biological and Pharmaceutical Bulletin 25(3): 361-366.

Lindh, M., A. Karlén and U. Norinder (2017). "Predicting the Rate of Skin Penetration Using an Aggregated Conformal Prediction Framework." Molecular Pharmaceutics 14(5): 1571-1576.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Luo, W., S. Medrek, J. Misra and G. J. Nohynek (2007). "Predicting human skin absorption of chemicals: Development of a novel quantitative structure activity relationship." Toxicology and Industrial Health 23(1): 39-45.

Luo, W., H. Nguyen, Q. Telesford and W. Fung (2004). "A New QSAR Model For Human Skin Absorption." Toxicologist 78(1-S): 324-325.

Magnusson, B. M., Y. G. Anissimov, S. E. Cross and M. S. Roberts (2004). "Molecular size as the main determinant of solute maximum flux across the skin." Journal of Investigative Dermatology 122(4): 993-999.

McKone, T. E. and R. A. Howd (1992). "Estimating dermal uptake of nonionic organic chemicals from water and soil: I. Unified fugacity-based models for risk assessments." Risk Anal 12(4): 543-557.

Milewski, M. and A. L. Stinchcomb (2012). "Estimation of maximum transdermal flux of nonionized xenobiotics from basic physicochemical determinants." Molecular Pharmaceutics 9(7): 2111-2120.

Mitragotri, S. (2002). "A theoretical analysis of permeation of small hydrophobic solutes across the stratum corneum based on Scaled Particle Theory." Journal of Pharmaceutical Sciences 91(3): 744-752.

Mitragotri, S. (2003). "Modeling skin permeability to hydrophilic and hydrophobic solutes based on four permeation pathways." Journal of Controlled Release 86(1): 69-92.

Moody, R. P. and H. MacPherson (2003). "Determination of dermal absorption QSAR/QSPRs by brute force regression: Multiparameter model development with molsuite 2000." Journal of Toxicology and Environmental Health - Part A 66(20): 1927-1942.

Moss, G. P. and M. T. D. Cronin (2002). "Quantitative structure-permeability relationships for percutaneous absorption: Re-analysis of steroid data." International Journal of Pharmaceutics 238(1-2): 105-109.

Neely, B. J., S. V. Madihally, R. L. Robinson Jr and K. A. M. Gasem (2009). "Nonlinear quantitative structure-property relationship modeling of skin permeation coefficient." Journal of Pharmaceutical Sciences 98(11): 4069-4084.

Neumann, D., O. Kohlbacher, C. Merkwirth and T. Lengauer (2006). "A fully computational model for predicting percutaneous drug absorption." Journal of Chemical Information and Modeling 46(1): 424-429.

Patel, H., W. T. Berge and M. T. D. Cronin (2002). "Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals." Chemosphere 48(6): 603-613.

Potts, R. O. and R. H. Guy (1992). "Predicting Skin Permeability." Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists 9(5): 663-669.

Potts, R. O. and R. H. Guy (1995). "A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity." Pharm Res 12(11): 1628-1633.

Poulin, P. and K. Krishnan (2001). "Molecular structure-based prediction of human abdominal skin permeability coefficients for several organic compounds." J Toxicol Environ Health A 62(3): 143-159.

Pugh, W. J., I. T. Degim and J. Hadgraft (2000). "Epidermal permeability-penetrant structure relationships: 4, QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge." International Journal of Pharmaceutics 197(1-2): 203-211.

Pugh, W. J. and J. Hadgraft (1994). "Ab initio prediction of human skin permeability coefficients." Int. J. Pharm. 103(Mar): 163-178.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Riviere, J. E. and J. D. Brooks (2005). "Predicting skin permeability from complex chemical mixtures." Toxicol Appl Pharmacol 208(2): 99-110.

Riviere, J. E. and J. D. Brooks (2007). "Prediction of dermal absorption from complex chemical mixtures: incorporation of vehicle effects and interactions into a QSPR framework." SAR QSAR Environ Res 18(1-2): 31-44.

Riviere, J. E. and J. D. Brooks (2011). "Predicting skin permeability from complex chemical mixtures: dependency of quantitative structure permeation relationships on biology of skin model used." Toxicol Sci 119(1): 224-232.

Riviere, J. E., J. D. Brooks, W. T. Collard, J. Deng, G. de Rose, S. P. Mahabir, D. A. Merritt and A. A. Marchiondo (2014). "Prediction of formulation effects on dermal absorption of topically applied ectoparasiticides dosed in vitro on canine and porcine skin using a mixture-adjusted quantitative structure permeability relationship." Journal of Veterinary Pharmacology and Therapeutics 37(5): 435-444.

Rocco, P., F. Cilurzo, P. Minghetti, G. Vistoli and A. Pedretti (2017). "Molecular Dynamics as a tool for in silico screening of skin permeability." Eur J Pharm Sci.

Samaras, E. G., J. E. Riviere and T. Ghafourian (2012). "The effect of formulations and experimental conditions on in vitro human skin permeation - Data from updated EDETOX database." International Journal of Pharmaceutics 434(1-2): 280-291.

Santos-Filho, O. A., A. J. Hopfinger and T. Zheng (2004). "Characterization of skin penetration processes of organic molecules using molecular similarity and QSAR analysis." Mol Pharm 1(6): 466-476.

Steinmetz, F. P., J. C. Madden and M. T. Cronin (2015). "Data Quality in the Human and Environmental Health Sciences: Using Statistical Confidence Scoring to Improve QSAR/QSPR Modeling." J Chem Inf Model 55(8): 1739-1746.

ten Berge, W. (2009). "A simple dermal absorption model: Derivation and application." Chemosphere 75(11): 1440-1445.

Thomas, J., S. Majumdar, S. Wasdo, A. Majumdar and K. B. Sloan (2007). "The effect of water solubility of solutes on their flux through human skin in vitro: An extended Flynn database fitted to the Roberts-Sloan equation." International Journal of Pharmaceutics 339(1-2): 157-167.

Walker, J. D., R. Rodford and G. Patlewicz (2003). "Quantitative structure-activity relationships for predicting percutaneous absorption rates." Environmental Toxicology and Chemistry 22(8): 1870-1884.

Williams, F. M., H. Rothe, G. Barrett, A. Chiodini, J. Whyte, M. T. D. Cronin, N. A. Monteiro-Riviere, J. Plautz, C. Roper, J. Westerhout, C. Yang and R. H. Guy (2016). "Assessing the safety of cosmetic chemicals: Consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern)." Regulatory Toxicology and Pharmacology 76: 174-186.

Xu, G., J. M. Hughes-Oliver, J. D. Brooks, J. L. Yeatts and R. E. Baynes (2013). "Selection of appropriate training and validation set chemicals for modelling dermal permeability by U-optimal design." SAR and QSAR in Environmental Research 24(2): 135-156.

Yu, Y. J., R. X. Su, L. B. Wang, W. Qi and Z. M. He (2010). "Comparative QSPR modeling of skin permeability for organic compounds using stepwise-MLR and PLS." Fresenius Environmental Bulletin 19(5): 832-837.

Zhang, K., M. H. Abraham and X. Liu (2017). "An equation for the prediction of human skin permeability of neutral molecules, ions and ionic species." Int J Pharm 521(1-2): 259-266.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Zhang, K., M. Chen, G. K. E. Scriba, M. H. Abraham, A. Fahr and X. Liu (2012). "Human skin permeation of neutral species and ionic species: Extended linear free-energy relationship analyses." Journal of Pharmaceutical Sciences 101(6): 2034-2044.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Results of the second stage of the scientific review

The detailed results of the critical appraisal of dermal absorption models are provided as Supporting Information (file name: "Model Review.xlsx"). This Excel file contains three tabs: one with the review of the single substance models, one with the review of the mixture models and one explaining the variable symbols used in the specified algorithms.

It should be noted that there is some redundancy in the variable symbols as different authors sometimes use different symbols. For easy reference to the original papers mostly the notation of their authors has been used. In some instances of often used symbols, it was decided to use a unique symbol, e.g. (log) K_{ow} for the octanol-water partitioning coefficient was always preferred over (log) P and k_p for the permeability constant was always preferred over P or other designations.

Below a short explanation is provided for the headings and subheadings of the columns in the review tabs, if not self-evident. If papers were excluded based on any question of the hierarchical second stage of the scientific review, the questions that follow thereafter were not answered. In these cases "n/a" (meaning not applicable) was entered in the respective fields in the Excel file.

Table D.4: Explanation of headings for the file "Model Review.xlsx" with supplementary information from the the scientific review

Heading	Subheading	Explanation
Paper #		Sequential number
LRT-ID		Unique ID of the paper used in the TNO Literature Review Tool (LRT)
LRT_link		Link that can be pasted in the address bar of the web-based LRT in order to navigate to the reviewed paper
Citation		Paper reference
Remark		
Model Label No.		Models appearing in the same paper are numbered in order of appearance. Default number is 1
1. Algorithm based model?	Yes/No	
	Reproducible	Preliminary evaluation based on presence of a well described algorithm for algorithm based models and on the availability of a detailed description of model structure and the complete training set with all input parameters for each substance for machine learning models
	Motivation	Justification, if deemed necessary
2. Addresses a clearly defined need in pesticide RA?	Yes/No	
	Specification	Reason for answering yes or no to question 2
	Chemical group(s)	Specifies chemical groups addressed in local models
3. More recent similar models?	Yes/No	
	Remark	Lists reference to most recent model if question 3 is answered with "Yes".
4. Details of the training set provided?	Yes/No	
	Chemical IDs of all members of the training set	Answered with "Yes", when these data are provided.
	Measured absorption parameters	Answered with "Yes", when these data are provided.
	Parameter values used in model	Answered with "Yes", when these data are provided.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Heading	Subheading	Explanation
	Other	Specifies any other relevant information provided on the dataset
5. Can the model be reproduced based the materials and methods of the paper?	Yes/No	
	Specification	Indicates whether the model can be completely or partly reproduced
	Motivation	When only partly reproducible, it is indicated here which part can and which part cannot be reproduced
6. Internal validation	Yes/No	
	R^2	Lists value if provided
	Adjusted R^2	Lists value if provided
	RMSE	Lists value if provided
	number of pre-dictors	Lists value if provided
	number in training set	Lists value if provided
	Other	Lists additional relevant value if provided
		Specified the type of additionally provided value
7. External validation	Yes/No	Lists value if provided
	R^2	Lists value if provided
	Adjusted R^2	Lists value if provided
	RMSE	Lists value if provided
	number of pre-dictors	Lists value if provided
	number in valida-tion set	Lists additional relevant value if provided
	Other	Specified the type of additionally provided value
		Lists value if provided
8. Mechanistic interpretation	Option	Four different options are provided, indicated with A, B, C or D (description provided in Table C.1)
	Motivation	Justification, if deemed necessary
9. Proprietary model?	Yes/No	
Model parameters		
Model algorithm(s)		

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix E – Data plausibility check and curation

For the purpose of this project, a column containing record identifiers 1-6842 relating to individual replicated of experiments was introduced to the EFSA dataset.

First, active substance SMILES codes were collected from the BfR internal IJC database or Pubchem. For cases where the no information could be found in neither sources, a google search was performed. This applied to the following active substance names in the provided EFSA dataset:

- SYN545192 was identified as Benzovindiflupyr
- for Aminopyralid-TIPA SMILES were found in ChemIDplus database
- XDE 777- XDE 779 / XDE-777 was identified as Fenpicoxamid and molecular weight of 614.648 g/mol was identical to molecular weight in EFSA dataset. The name of the a.s. remained XDE 777- XDE 779 / XDE-777
- Imidazole/BAS 590 02 F was found to be present in pesticide product Octave 46 WP which could be associated with Prochloraz manganese chloride complex also having the same molecular weight (1632.525 g/mol).

For several active substances, some additional information is reported below:

- According to the EFSA conclusion on the active substance Spinetoram, Spinetoram (XDE-175) is a mixture of two main components, XDE-175-J-major factor and XDE-175-L-minor factor. SMILES and consequently physico-chemical values and other parameters were calculated for XDE-175-J, which represents the major factor with an amount of 50-90% in the Spinetoram mixture.
- Beta Cyfluthrin and Cyfluthrin differ in isomeric composition. Thus, distinct active substance names were retained although no difference was made at the structural level (SMILES codes). Thus, physico-chemical properties are essentially identical because they were computed based on SMILES codes.
- Dimethenamid and Dimethenamid-P also differ in isomeric composition. Nevertheless, as no difference was made at the structural level (SMILES code), calculated descriptors are identical. For records 1677-1694 with the a.s. name Dimethenamid, the identity could be confirmed from the study report. For records 1695-1704, the tested active substance was identified as Dimethenamid, rather than the reported Dimethenamid-P and a correction was performed. For all other records, study reports were not available and the identity was kept as Dimethenamid P.
- Mancozeb was excluded from the dataset due to ambiguous structure. Referring to the DAR Part B.1, mancozeb is defined as a bulk chemical composition and there is no unique chemical molecule which may be identified as mancozeb.
- In case of AE 1801486, no information on active substance chemical name could be found in internal regulatory documents or internet search.
- Entries for Mancozeb and AE 1801486 were kept in the dataset but were completely excluded from the evaluation.They are to be filtered out by selecting "Y" at the new column "Include?" in the datafile provided as Supporting Information (file name: ”Human *in vitro* PPP EFSA dataset with add parameters and predictions.xlsx”).

Based on structure (SMILES code), the molecular weight (MW) was computed with IJC. This served as plausibility check for identity of active substance and correctness of the SMILES codes. Computed and reported MWs were compared. Two out of 191 listed active substances showed deviating ratios:

- Flupyrsulfuron methyl showed a difference of computed (465.36) versus reported MW (487.40) of 22.04 g/mol, which might account for sodium. When referring to the original study report, the tested active substance was Flupyrsulfuron methyl monosodium salt. Thus, reported MW is correct as reported in the dermal absorption study.
- The deviation recognised for the active substance Prochloraz-copper can be explained with giving the MW for the monomeric Prochloraz (376.70) in the EFSA dataset. The computed MW

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

(1641.09) related to the Prochloraz-copper complex, because it is based on SMILES codes which were in turn generated based on given active substance name. It can be assumed that monomeric Prochloraz will be absorbed and be the molecule of interest. Additionally, some of the physico-chemical descriptors could only be predicted for monomers (here: Prochloraz). Accordingly, the structure of Prochloraz was adopted.

In addition, a Pivot Table was created to check for different values for one a.s. name. Pivot and conditional formatting were used for highlighting duplicate values on MW to check for similar/same a.s.. Pivot and conditional formatting were further used for highlighting duplicate values on both rows to check whether similar a.s. are indeed the same despite differences in names. A line-by-line visual inspection was performed to identify missing or contradictory entries. If needed, study reports were consulted when available. The following substances were renamed without changes to the data to resolve duplicate entries for identical structures as well as for consistency, transparency and correctness:

- 2,4-D to 2,4-D (2,4-dichlorophenoxyacetic acid)
- 2,4-D EHE to 2,4-D EHE (2-ethylhexyl 2-(2,4-dichlorophenoxy)acetate)
- Cymoxynil to Cymoxanil
- Cyflometofen to Cyflumetofen
- Diruon to Diuron
- E2Y to Chlorantraniliprole
- Imidazole/BAS59002F, identified as Prochloraz manganese, to Imidazole/BAS 590 02 F (Prochloraz-manganese)
- Oxyfluropfen to Oxyfluorfen
- SYN520453 (Isopyrazam) was renamed to Isopyrazam.
- XDE-729 methyl, identified as Halauxifen-methyl, to XDE-729 methyl (Halauxifen-methyl)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix F – Additional information on the collection of physical-chemical data

Calculation of hydrogen bond acceptor activity and hydrogen bond donor activity: additional considerations

Although superseded by Model003, Model001 was used for prediction of Hydrogen bond donor/acceptor activities, as it was thought to most closely reflect the tool used by the authors of the skin absorption model where A (there: Hd) and B (there: Ha). In addition, the webservice based on Model003 did not work reliably. Moreover, there are a few limitations reported for Model003 at an additional webpage referring to the applicability domain mentioned by the developers. First, for compounds outside the applicability domain errors can be the result of prediction. This might be the cases for substances with XLogP < -0.03, MLogP < 1.63 AND TpSA > 31.8, the compound contains a phenol or hetero atom and if a 5-ring is connected to a general functional group by a single bond. In addition, it is highlighted by the developers that A, the hydrogen bond acidity, is the most difficult Abraham descriptor to model. For that reason the value for A might be predicted negative, but usually A should be zero when the number of hydrogen bond donors is zero. Regarding these limitations, the application of Model001 was regarded as acceptable within the scope of the project.

When entering the required input data for the computation with Model001 the output comprises five Abraham descriptors E, S, A, B, V. Defining A as the solute overall hydrogen bond acidity which relates to hydrogen bond donor activity and B as the solute overall hydrogen bond basicity which relates to hydrogen bond acceptor activity. For Model001 and the underlying training set, adjusted R^2 values of 0.6262 and 0.8327 were stated for A and B, respectively.

Calculation of EHOMO and ELUMO energy values for Fu 2002: Reproducibility of MOPAC2016

For the training set used by the authors of models 4 and 5 (Fu, Ma and Liang , 2002), the AM1 method was employed for calculation of these values. Accordingly, the AM1 method was also chosen to predict EHOMO and ELUMO values for the pesticidal active substances. It was checked whether MOPAC2016 can reproduce the EHOMO and ELUMO values calculated by Fu, Ma and Liang (2002) to a satisfactory degree. SMILES codes for were obtained for 12 substances selected from the training set used by Fu *et al.* (2002) from PubChem public database. The results are summarized in

Table F.1.For testosterone geometry optimization by MOPAC2016 failed and EHOMO/ELUMO values were not calculated. For dexamethasone, OpenBabel failed to produce a viable 3D structure as atom 20 and atom 25 were superimposed, producing an error message in MOPAC2016. HOMO/LUMO energies as reported by Fu *et al.* (2002) and those calculated by the method described above were not fully identical. Despite application of the same method (AM1), values differed by between 0.0006 and 1.9366 eV with a median of 0.2548 eV (n = 20 pairs). The largest difference between predictions by Fu *et al.* (2002) and those performed in this project were observed for ethylbenzene and 2-naphthol. It can only be speculated that observed differences may be related to the implementation of the AM1 method in MOPAC2016, as the tool used by Fu *et al.* (2002) is not known. Nevertheless, for the purpose of this project, the level of reproduction was considered sufficient. This step in the procedure should be taken into account as a potential source of error when assessing predictivity of the model against the EFSA dataset and when applying the model for other purposes.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table F.1: Comparison of reported EHOMO / ELUMO values for the training data set used by Fu, Ma and Liang (2002) to values predicted with the methodology described here.

Substance	SMILES	EHOMO [eV]		ELUMO [eV]	
		MOPAC	Fu *et al.*	MOPAC	Fu *et al.*
diethylether	CCOCC	-10.392	-10.3931	2.981	2.9816
butanoic acid	CCCC(=O)O	-11.247	-11.5020	1.058	1.0285
methanol	CO	-11.134	-11.1349	3.780	3.7783
1-hexanol	CCCCCCO	-10.934	-10.8475	3.463	3.3692
1-decanol	CCCCCCCCCCO	-10.909	-10.8493	3.392	3.3320
ethylbenzene	CCC1=CC=CC=C1	-8.015	-9.2984	-1.399	0.5376
3,4-di-methylphenol	CC1=C(C=C(C=C1)O)C	-8.586	-8.8561	-0.506	0.4559
4-chloro-phenol	C1=CC(=CC=C1O)Cl	-8.870	-9.1246	-0.329	0.0947
4-bromo-phenol	C1=CC(=CC=C1O)Br	-8.931	-9.1892	-0.395	0.0204
2-naphthol	C1=CC=C2C=C(C=CC2=C1)O	-7.712	-8.5697	-1.858	-0.3443
teststerone	CC12CCC3C(C1CCC2O)CCC4=CC(=O)CCC34C	n.d.	-10.0129	n.d.	0.0143
dexamethasone	CC1CC2C3CCC4=CC(=O)C=CC4(C3(C(CC2(C1(C(=O)CO)O)C)O)F)C	n.d.	-10.1421	n.d.	-0.4457

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix G – Models' applicability and codes

Model	Model sub-version code	Model sub-version	Total excluded (percentage of whole dataset)	Exclusion criteria - presented in the order applied (no of replicates excluded)
1 (all sub-versions)	Model 1, t_lag=0 OR Model 1, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**580 (8.5 %)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517)
2 (all sub-versions)	Model 2, t_lag=0 OR Model 2, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**580 (8.5 %)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517)
3 (sub-versions calculated with computed Saq values and both lag time options)	Model 3 (calc. S_aq), t_lag=0 OR Model 3 (calc. S_aq), calculated t_lag	2 sub-versions (out of 4) calculated with: i. calculated Saq values for J_{max}→k_p conversion ii. different tlag assumptions for k_p →%DA transformation (tlag= 0 OR tlag=calculated)	**580 (8.5 %)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517)
3 (sub-versions calculated with experimental Saq values and both lag time options)	Model 3 (calc. S_aq), t_lag=0 OR Model 3 (calc. S_aq), calculated t_lag	2 sub-versions (out of 4) calculated with: i. experimental Saq values for J_{max}→k_p conversion ii. different tlag assumptions for k_p →%DA transformation (tlag= 0 OR tlag=calculated)	**762 (11.2%)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517) iv. additional criterion: absence of experimental data on water solubility (149) v. additional criterion: experimental water solubility data=0·division is mathematically not allowed (33)
4 (all sub-versions)	Model 4, t_lag=0 OR Model 4, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**588 (8.6%)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. eHOMO eLUMO values computation not possible (14) iii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iv. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (511)
5 (all sub-	Model 5, t_lag=0	Both sub-versions with different tlag assumptions	**612 (9.0%)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Model	Model sub-version code	Model sub-version	Total excluded (percentage of whole dataset)	Exclusion criteria - presented in the order applied (no of replicates excluded)
versions)	OR Model 5, calculated t_lag	for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)		provided by EFSA (6) ii. eHOMO eLUMO values computation not possible (14) iii. Aminopyralid-Olamine excluded (24) iv. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) v. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (511)
6 (sub-versions calculated with computed Saq and both MP and lag time options)	Model 6 (calc.Saq, n.T), t_lag=0 OR Model 6 (calc.Saq), t_lag=0 OR Model 6 (calc.Saq, n.T), calculated t_lag OR Model 6 (calc.Saq), calculated t_lag	4 subversions (out of 8) calculated with: i. calculated Saq values for J_{max}→ k_p conversion ii. MP for liquids set at 25°C (marked as "nT") OR MP for liquids set as recorded in the database (no additional marking) iii. different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**580 (8.5%)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517)
6 (sub-versions calculated with experimental Saq and both MP and lag time options)	Model 6 (exp.Saq, n.T), t_lag=0 OR Model 6 (exp.Saq), t_lag=0 OR Model 6 (exp.Saq, n.T), calculated t_lag OR Model 6 (exp.Saq), calculated t_lag	4 subversions (out of 8) calculated with: i. experimental Saq values for J_{max}→ k_p conversion ii. MP for liquids set at 25°C (marked as "nT") OR MP for liquids set as recorded in the database (no additional marking) iii. different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**762 (11.2%)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517) iv. additional criterion: absence of experimental data on water solubility (149) v. additional criterion: experimental water solubility data=0·division is mathematically not allowed (33)
7	Model 7	No sub-versions	**5250 (77%)**	Model applied only to a restricted dataset (see 2.7.1.7)
8 (all sub-versions)	Model 8, t_lag=0 OR Model 8, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**580 (8.5%)**	i. non-plausible negative experimental % Dermal Absorption (%DA) values as provided by EFSA (6) ii. implementation not possible: absence of data on the experimentally applied concentration as provided by EFSA (57) iii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items (517)

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Model	Model sub-version code	Model sub-version	Total excluded (percentage of whole dataset)	Exclusion criteria - presented in the order applied (no of replicates excluded)
9 (all sub-versions)	Model 9, t_lag=0 OR Model 9, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**5863 (86%)**	i. implementation not possible: no information on product co-formulants retrievable → mixture model not applicable ii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items
10 (all sub-versions)	Model 10, t_lag=0 OR Model 10, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**5863 (86%)**	i. implementation not possible: no information on product co-formulants retrievable → mixture model not applicable ii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items
11 (all sub-versions)	Model 11, t_lag=0 OR Model 11, calculated t_lag	Both sub-versions with different tlag assumptions for k_p→ %DA transformation (tlag= 0 OR tlag=calculated)	**5863 (86%)**	i. implementation not possible: no information on product co-formulants retrievable → mixture model not applicable ii. implementation not possible: non-applicability of the DAME model for non-liquid experimental items

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix H – Adjusted logit

Many of the plots in the report use either a logarithmic scale or an "adjusted logit scale" on one or both axes. This is done always to try to enhance the communication and extraction of information. Logarithmic scales are familiar to scientists and need no explanation. The adjusted logit scale used here was introduced in the EFSA (2017) guidance on dermal absorption and some support for its use may be found in Annex B therein. Some additional description and justification is provide here.

The logit of a fraction f (or a percentage considered as a number between 0 and 1) is

$$\text{logit}(f) = \log [f/(1-f)].$$

Fractions between 0 and 1 are mapped/transformed to numbers which lie anywhere on the real line. The fraction 50% = 0.5 is mapped to 0, fractions above 50% to positive numbers and fractions below 50% to negative numbers. The effect of the transformation is to stretch out the scale near 0 and 1. In effect for fractions near 0, the logit is equivalent to the natural logarithm of the fraction and there is a similar property near 1. A logit scale is a good choice if one wishes to stretch out fractions near 0 and 1 in plots of data. It is also a good choice if the precision of measurement is higher near 0 and 1, for example if the typical error/variability of measurement of the fraction is proportional to the fraction being measured for small fractions.

The adjusted logit scale used in EFSA (2017) was used to address the fact that some measurements of absorption in the database are actually 0% and such measurements would have to be excluded if using a pure logit scale. In the adjusted logit scale, fractions are first "shrunk" a little towards 50% before calculating the logit, i.e. by calculating

$$\ell = \text{adjusted logit}(f) = \text{logit}(0.5 + 0.9995(f-0.5)) \qquad (1)$$

We can invert this, i.e. find f from ℓ, by first undoing the logit:

$$f_{shrunk} = e^{\ell} / (e^{\ell} + 1) \qquad (2)$$

and then undoing the "shrinking":

$$f = 0.5 + (f_{shrunk} - 0.5)/0.9995 \qquad (3)$$

The corresponding "adjusted logit scale" is used in some figures which follow but the axis tick marks show the original percentage absorption to facilitate interpretation of figures. The adjusted logit scale is also used for averaging replicates as follows:

First calculate the adjusted logit of each replicate using equation (1), then calculate the arithmetic mean of the adjusted logit values and finally calculate the inverse of the arithmetic mean using (2) and (3). The result is an "adjusted-logit mean" in the same sense that using the process with logarithms instead of logits gives the geometric mean.

The key to use of the adjusted logit scale is the precision of measurement of absorption and how it depends on the actual magnitude of absorption. This can be assessed, at least partially, by looking at the difference between replicate absorption measurement and assessing how the typical magnitude of the difference depends on the level of absorption being measured. This is illustrated in the figures below:

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

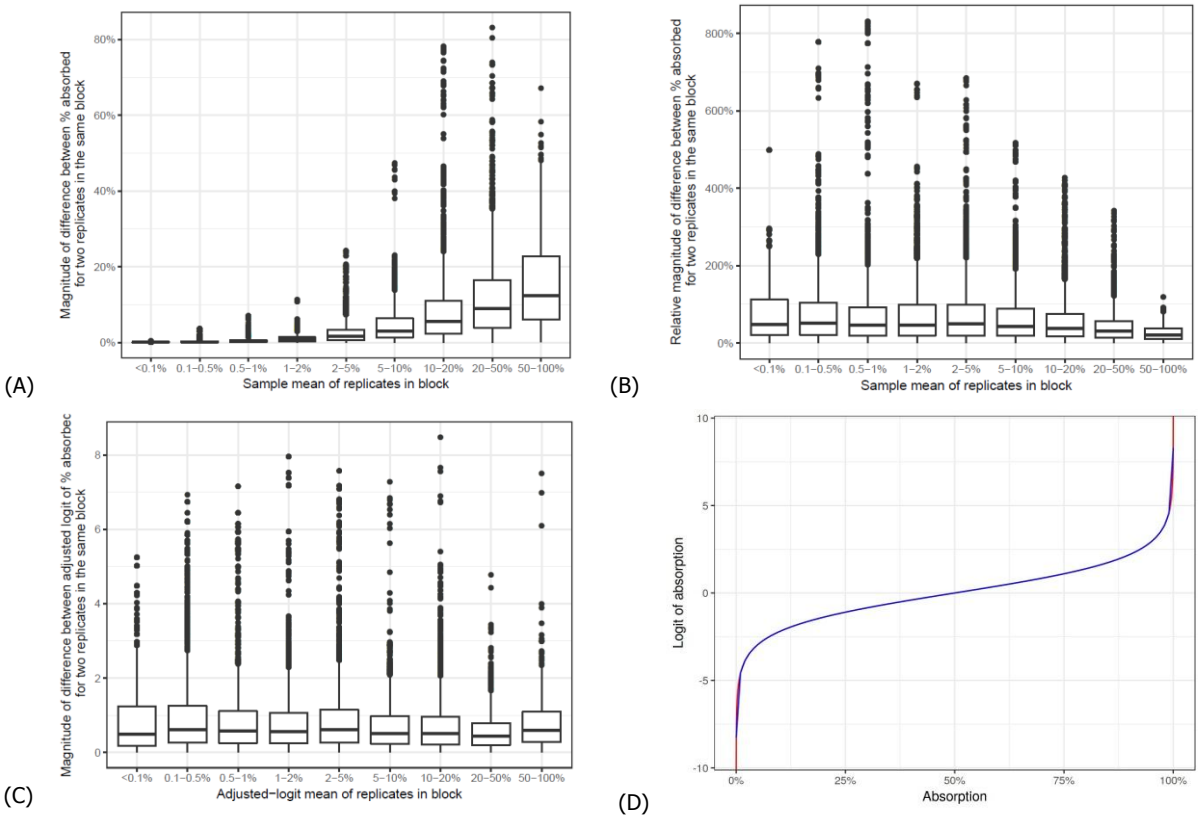


Figure H.1: Illustration of scales used. (A) Untransformed (raw): Dependence of the (absolute) magnitude of the difference between two replicate measurements of absorption on the magnitude of absorption (B) Relative: Dependence of the (relative) magnitude of the difference between two replicate measurements of absorption on the magnitude of absorption (C) Transformed: Dependence of the (absolute) magnitude of the difference between two replicate measurements of absorption on the magnitude of absorption, with adjusted-logit (replicate/mean) values for %DA (D) the logit (red) and adjusted logit (blue) transformations as functions of absorption.

Figure H.1(A) shows that there is a strong relationship between the magnitude of variation between replicates and the absorption level and suggests that some alternative scale should be considered when analyzing dermal absorption data. A common solution is to use logarithms but Figure H.1 (B) shows that the relative magnitude of variation is not constant but decreases near the upper hand of the absorption scale. Figure H.1 (C) shows that the adjusted logit is largely successful in making the magnitude of variation between replicates homogeneous across the absorption scale. Note that the difference of each pair of replicates from the same experiment is shown: this gives more weight to experiments having more replicates. Differences between pairs of replicates were used rather than the difference between a replicate and the sample mean for the experiment because the difference between pairs isolate the intra-experiment variation.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix I – Bayesian Random Effects Model (BREM)

The Bayesian Random Effects model used to analyse the predictive performance of model 7 has a simple structure. The adjusted-logit values of absorption measurements are denoted y_i and the base 10 logarithms of model 7 predicted absorptions by x_i. Then

$$y_i = \alpha + \beta\, x_i + \gamma_{AS[i]} + \delta_{block[i]} + \sigma\varepsilon_i$$

Here α is the intercept for a linear relationship and β is the slope. $\gamma_{AS[i]}$ is a random effect for the active substance involved for observation i and $\delta_{block[i]}$is a random effect for the block of replicates to which the observation belongs. $\sigma\varepsilon_i$ is the replicate variation where σ controls the scale of the variation and ε_i is assumed to be t-distributed with an uncertain number of degrees of freedom, to be inferred from the data. Each of the two kinds of random effect is assumed to be normally distributed with mean 0 and has its own standard deviation.

Inference for all parameters is Bayesian using Markov Chain Monte Carlo, implemented using the rjags package for R (Plummer, 2016) and using the following approach to specifying prior distributions for all models considered: fixed effects were assigned uniform distributions on the range from -20 to 20, standard deviations of fixed effects and the scale parameter for replicate variation were assigned uniform distributions on the range from 0 to 10 and degrees of freedom parameters were assigned a prior which was uniform on the range from 0 to 1 for the reciprocal of the degrees of freedom. The ranges for the uniform distributions for fixed effects and standard deviation and scale parameters are sufficiently wide that these are effectively flat priors on all possible values. The prior for degrees of freedom is Jeffrey's prior for a single parameter .The model described above was arrived at following the investigation of a number of related similar models:

- The t-distribution model for replicate variation was selected having considered also three other models: a normal error model and two models adding components for within-donor and between-donor variability; one with t-distributions for both components and one with a t-distribution for within-donor variability and a normal distribution for between-donor variability. Inference for all models was carried out using rjags and the selected model was chosen on the basis of comparison of Deviance Information Criterion values.
- The choice to include just the linear regression fixed-effect for model 7 and random effects for active substance and replicate block was also made on the basis of DIC having considered a number of other models. In this case, two other models had very similar DIC value but were more complex. The models considered were:
    a) The model above without a random effect for active substance.
    b) The model above with an additional fixed effect depending on formulation group
    c) The model above with an additional fixed effect for concentrate/dilution
    d) Model c) with an additional per-substance random effect for concentrate/dilution.
    e) Model c) with an additional fixed effect depending on formulation group.
    f) Model d) with an additional fixed effect depending on formulation group.

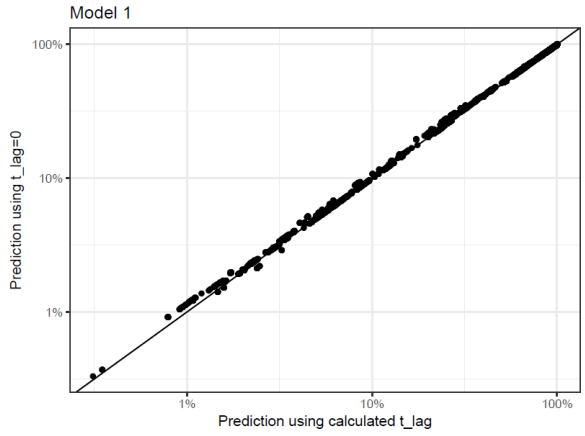
The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

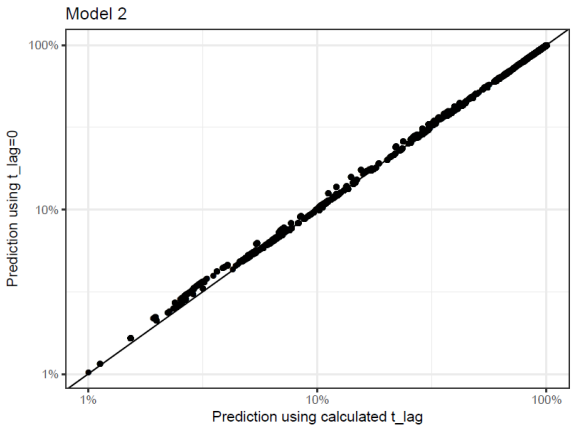
Appendix J – Influence of lag time: conversion of k_p / J_{max} into percentage absorption

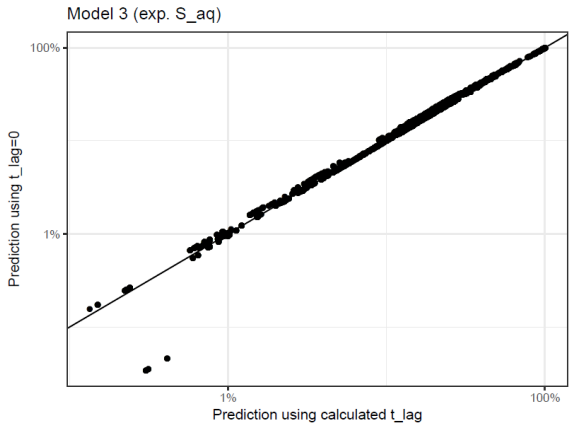
In order to convert the predicted permeability constant k_p or maximum flux J_{max}, selected models were combined with DAME as described in 2.7. This applies to all selected models with the exception of model 7. Two types of predictions were made for each model, one based on a lag time set to 0 ("worst case"), and one with a calculated lag time, predicted as described in 2.5.1.

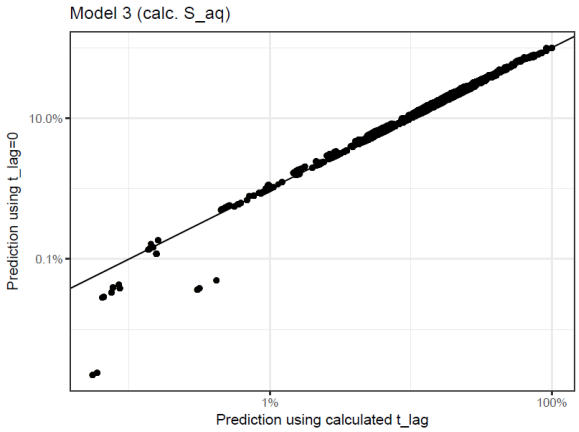
The following plots show the relationship between predictions using the two sub-versions (calculated tlag and tlag = 0) of each model or model version. Logarithmic scales are used on both axes in order to distinguish between different magnitudes of predicted absorption. A solid line highlights where the two versions make exactly the same predictions.

In the context of the further analyses performed - in particular on the correlation between measured and predicted values-, the differences between the two sub-versions of each model were not sufficient to merit producing a separate in-depth evaluation for each sub-version. When applicable, the calculated tlag sub-versions were used for all models in further assessment.

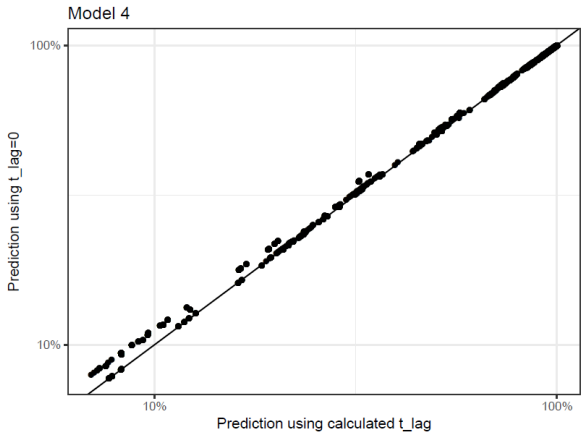


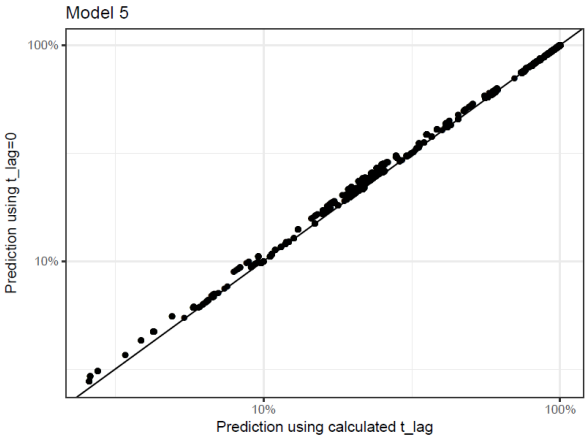


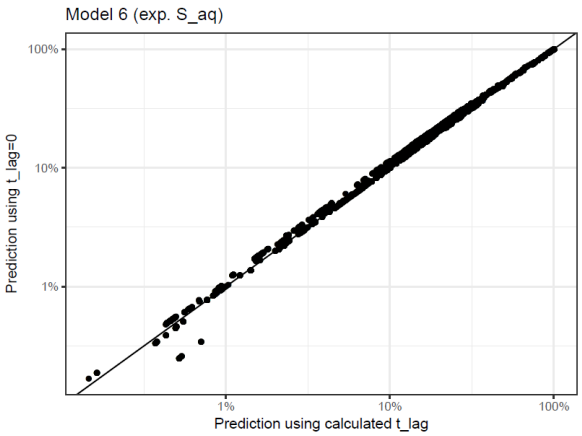


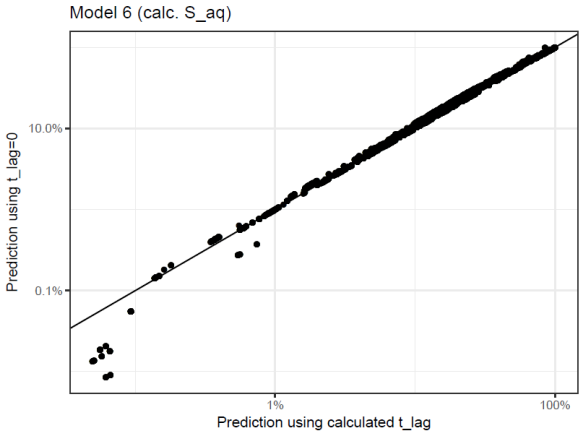


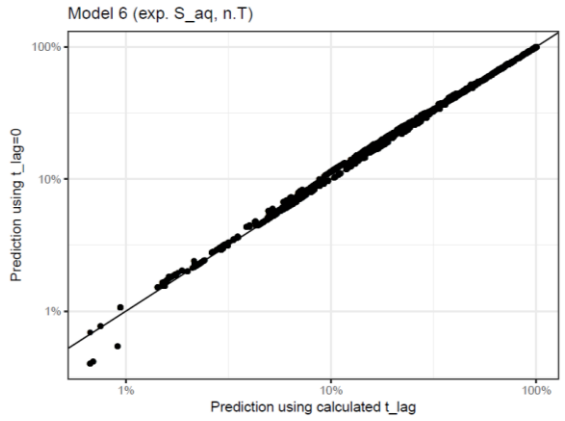
The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

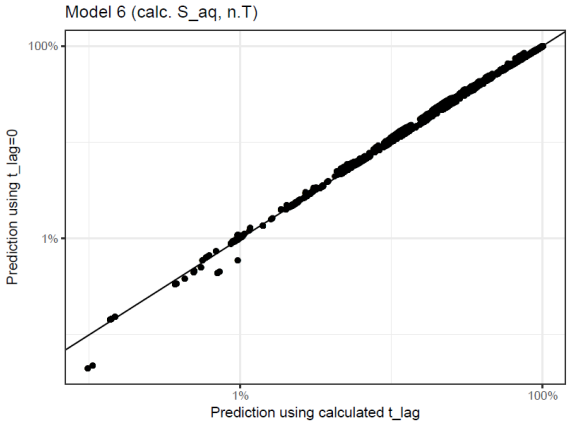












The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

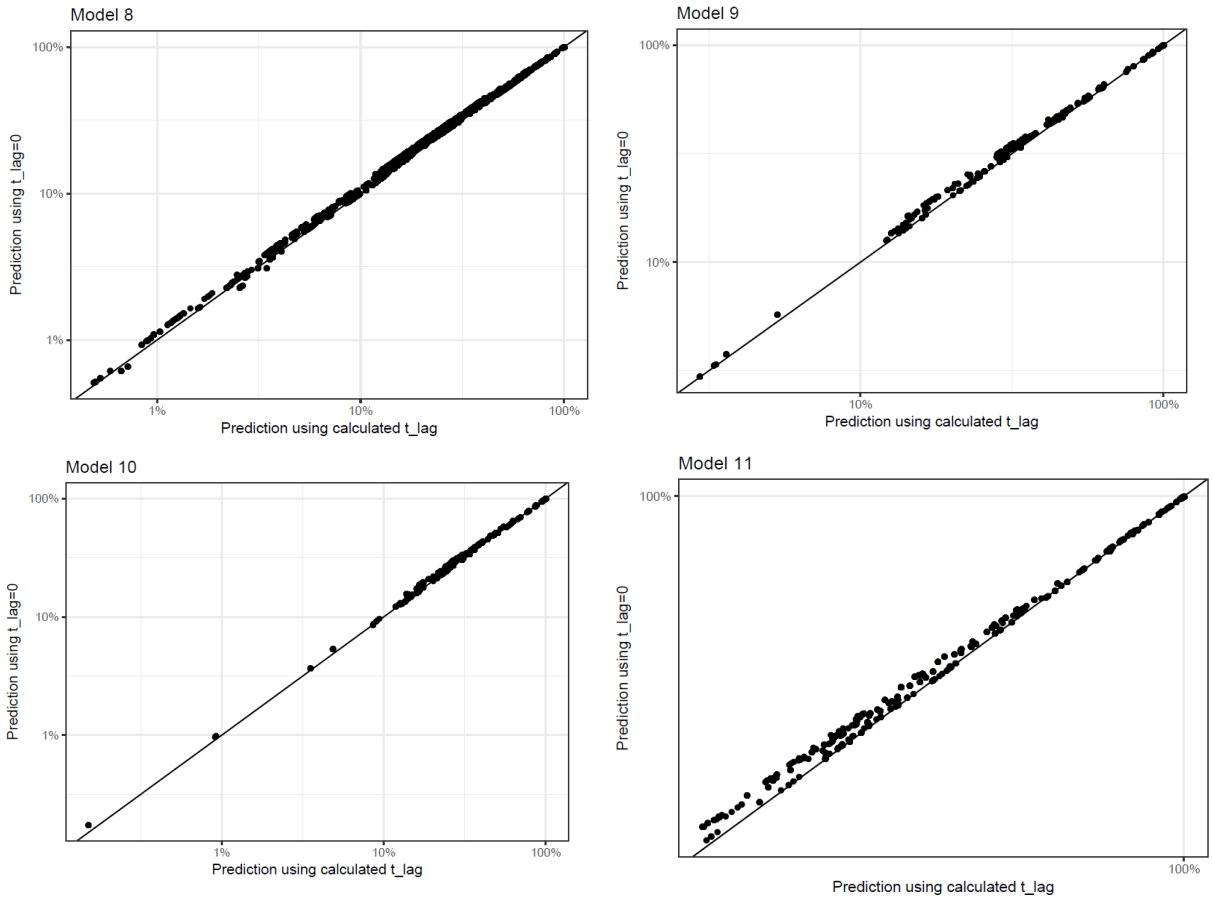
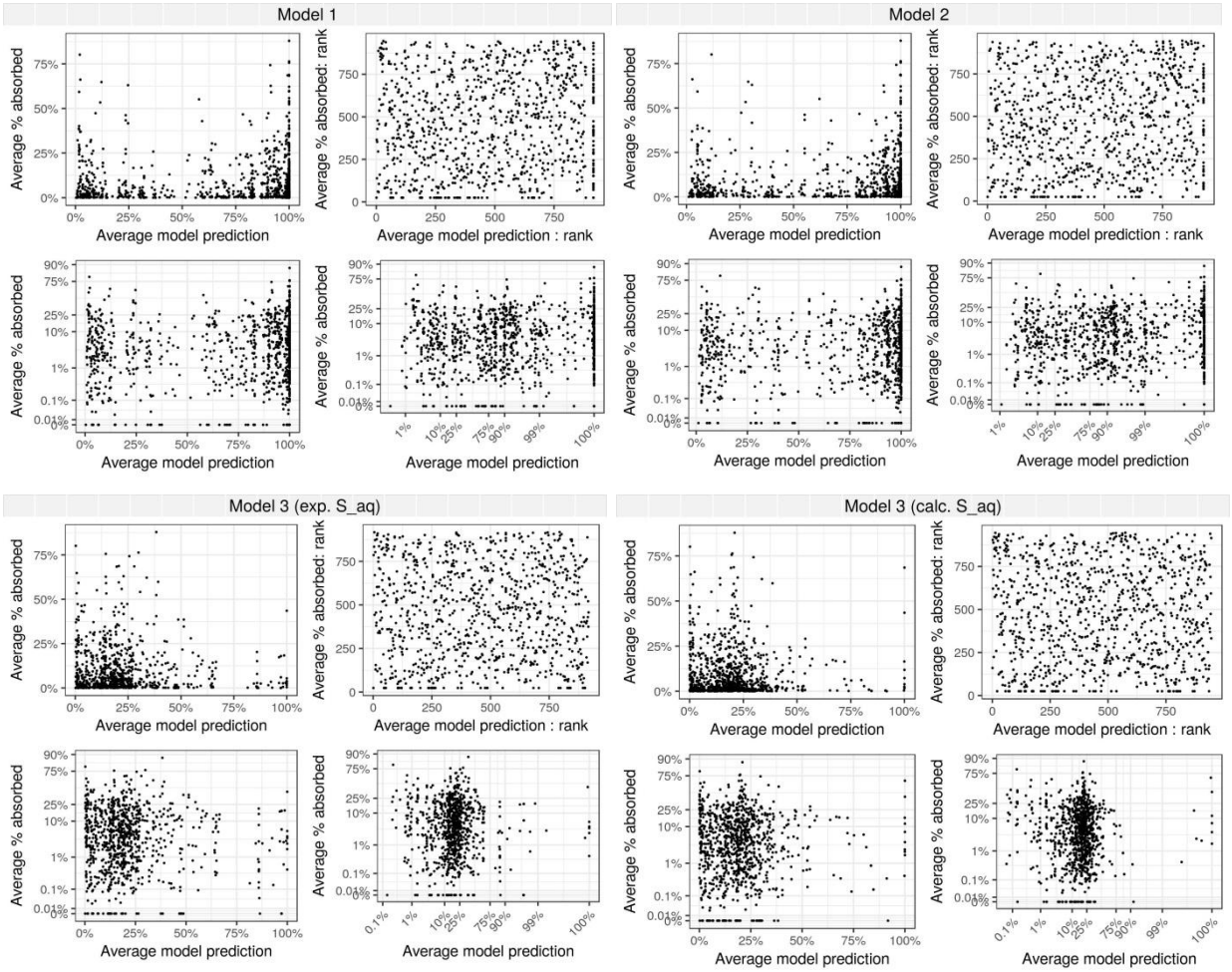


Figure J.1: Relationship between the predicted %DA values of each model (or model version) derived by either using tlag=0 or calculated tlag in the k_p-%DA transformation formula (Buist 2010 (DAME)). Model 7 is not depicted here since it directly yields %DA results. For models' explanations please refer to Appendix G.

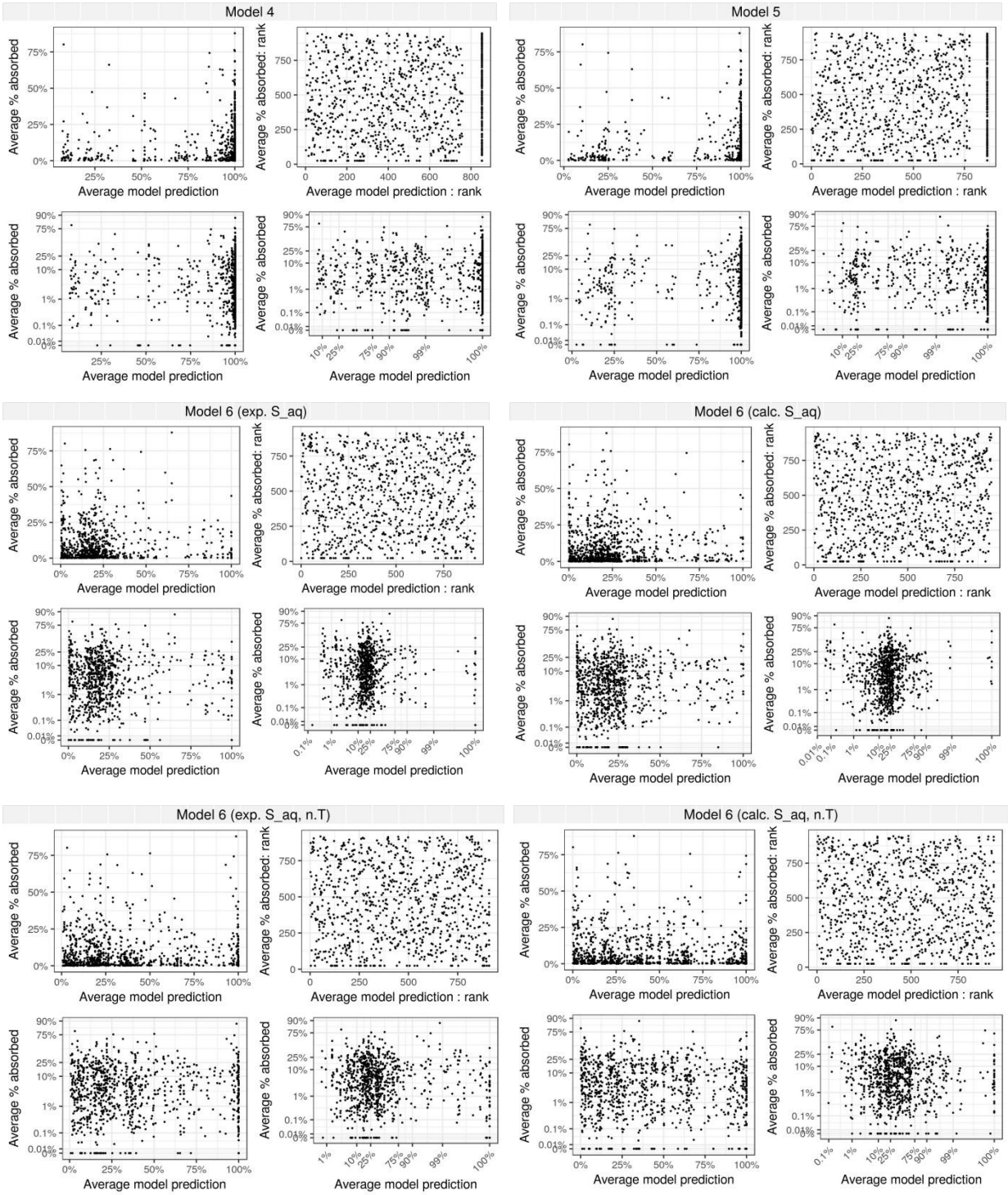
The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix K – Model predictions vs. average absorption measurements

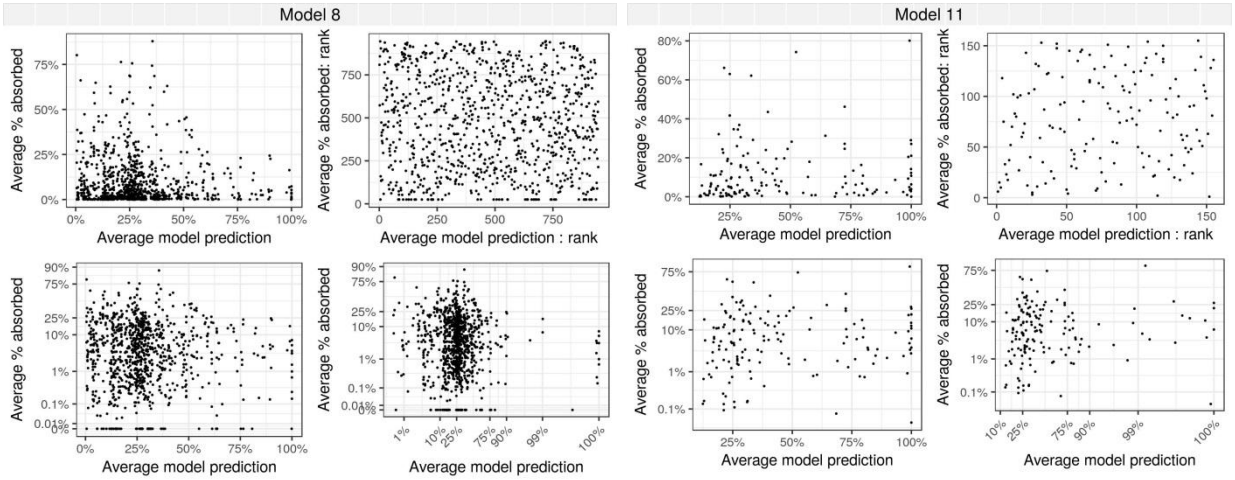
Below are presented the relationships between averages of replicates within blocks of replicates and corresponding average predictions; for models 1, 2, 3, 4, 5, 6, 8, 11 (and their subversions). For all these figures the calculated lag time version of the model was used (for explanations refer to 2.5.1 & Appendix J). Each figure shows 4 versions of the same relationship but using different scales on the axes. In each case the measured absorption is on the vertical axis and the absorption predicted by the model is on the horizontal axis. The top-left panel shows the relationship using the unmodified percentage absorbed scale for both axes. The bottom-left panel uses the adjusted logit scale for the vertical axis. The bottom-right panel uses the adjusted logit scale for both axes. The top-right panel uses a rank scale for both axes. In other words, each measurement is replaced by its rank amongst all measurements and each prediction is replaced by its rank amongst all the predictions from the model. For models' explanations please refer to Appendix G for explanations on the logit axes transformation please refer to Appendix H.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

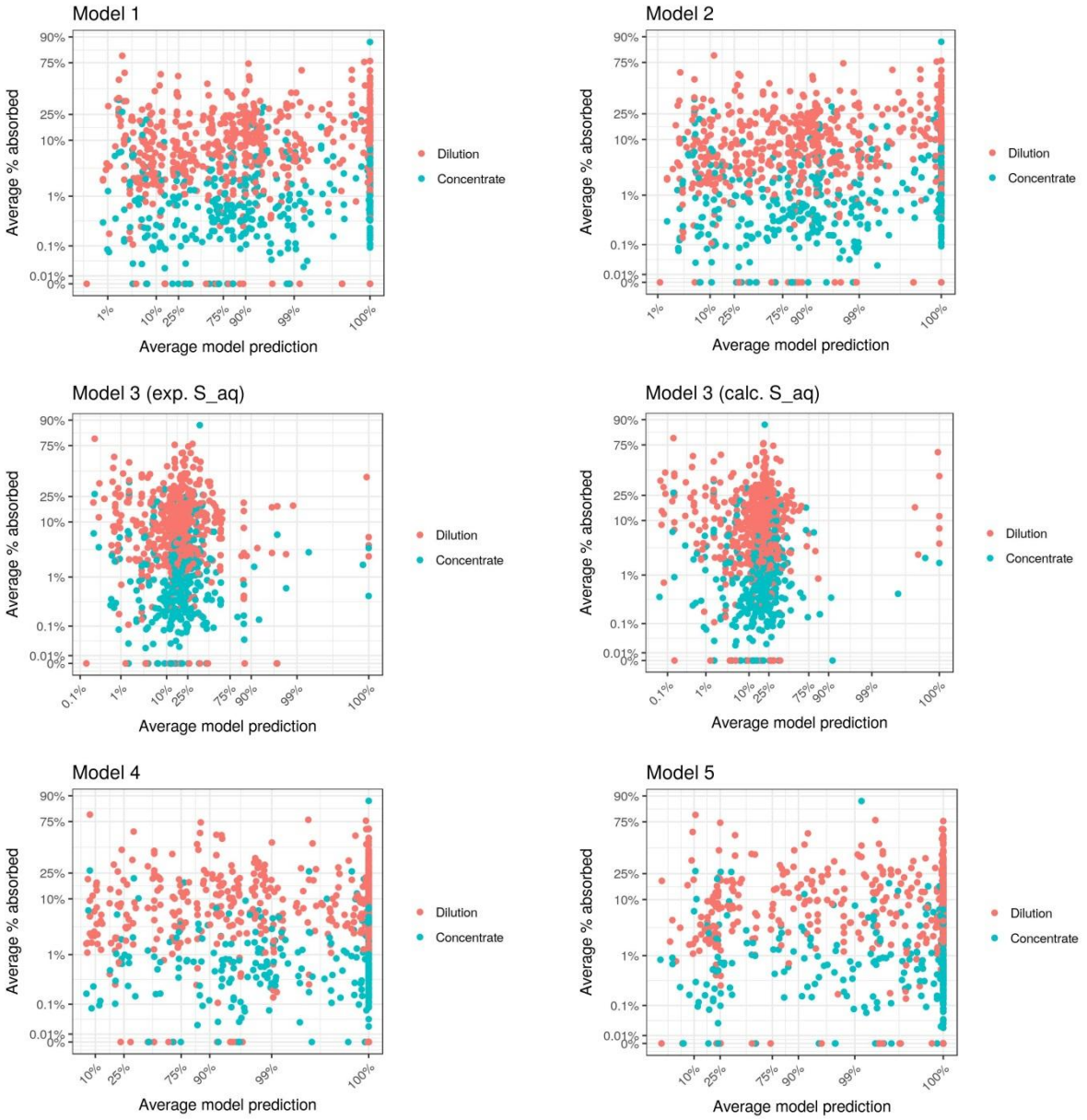


The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Appendix L – Influence of covariates

Below, the influence of sub-categorisation by concentration status (Figure L.1) and formulation type (Figure L.2) is depicted graphically. Models 7, 9 and 10 are presented in the main document. For explanations on the model's coding refer to Appendix G.

Concentrates versus in-use dilutions (Concentration status)



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

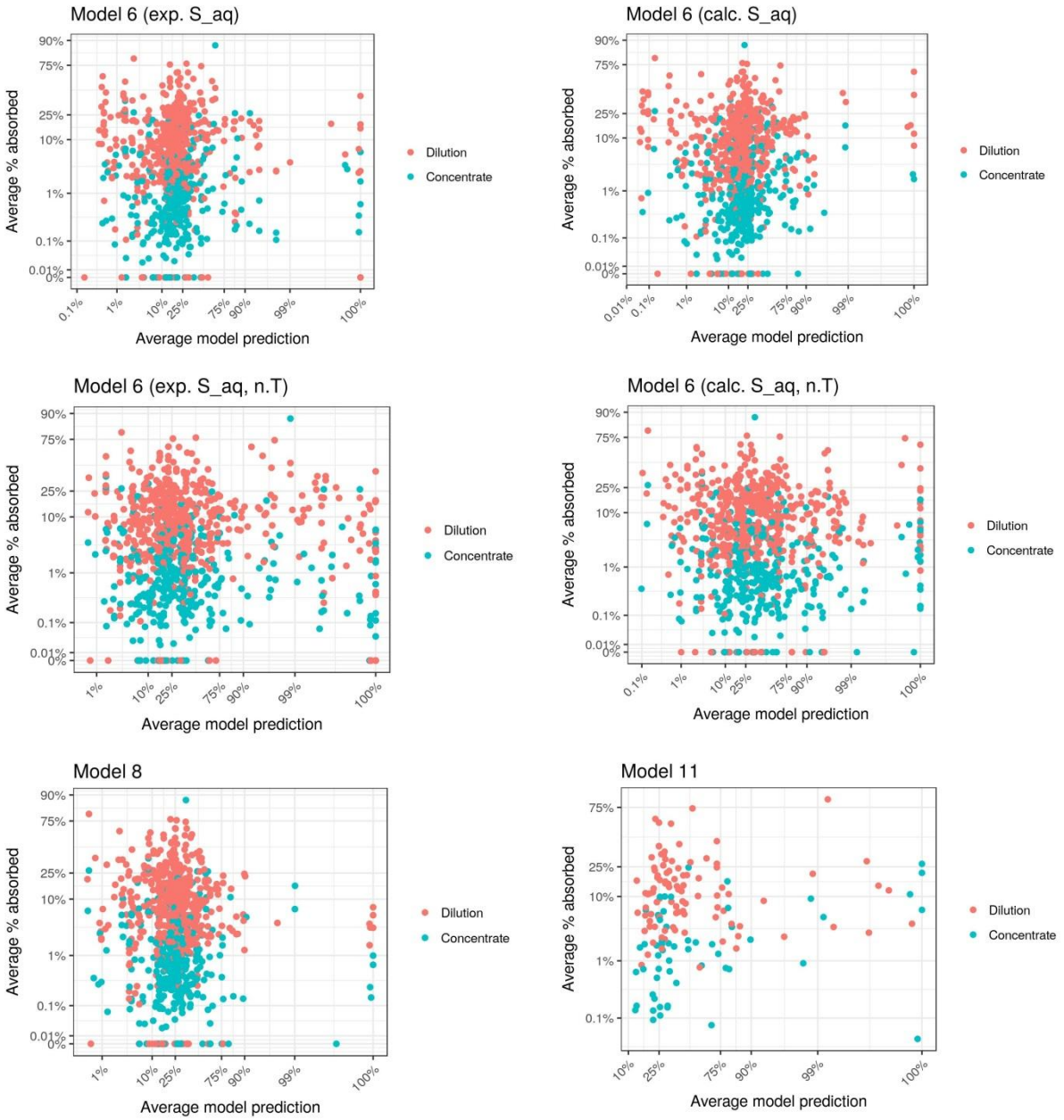
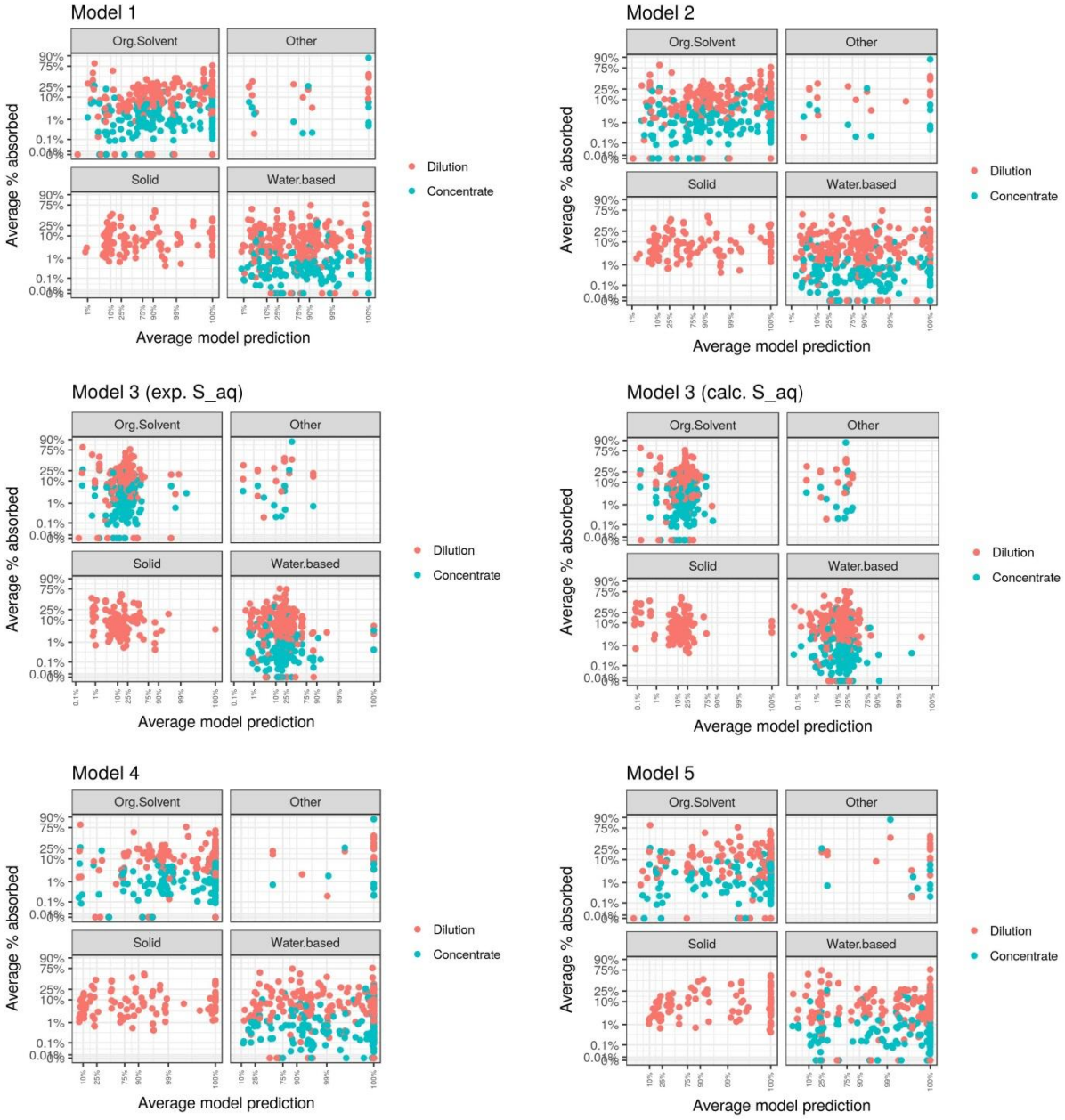


Figure L.1: Relationships between the averaged experimental % Dermal Absorption (%DA) values and the predicted ones divided into concentrates and dilutions. For models' explanations please refer to Appendix G.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Merged formulation type and dilution status



The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

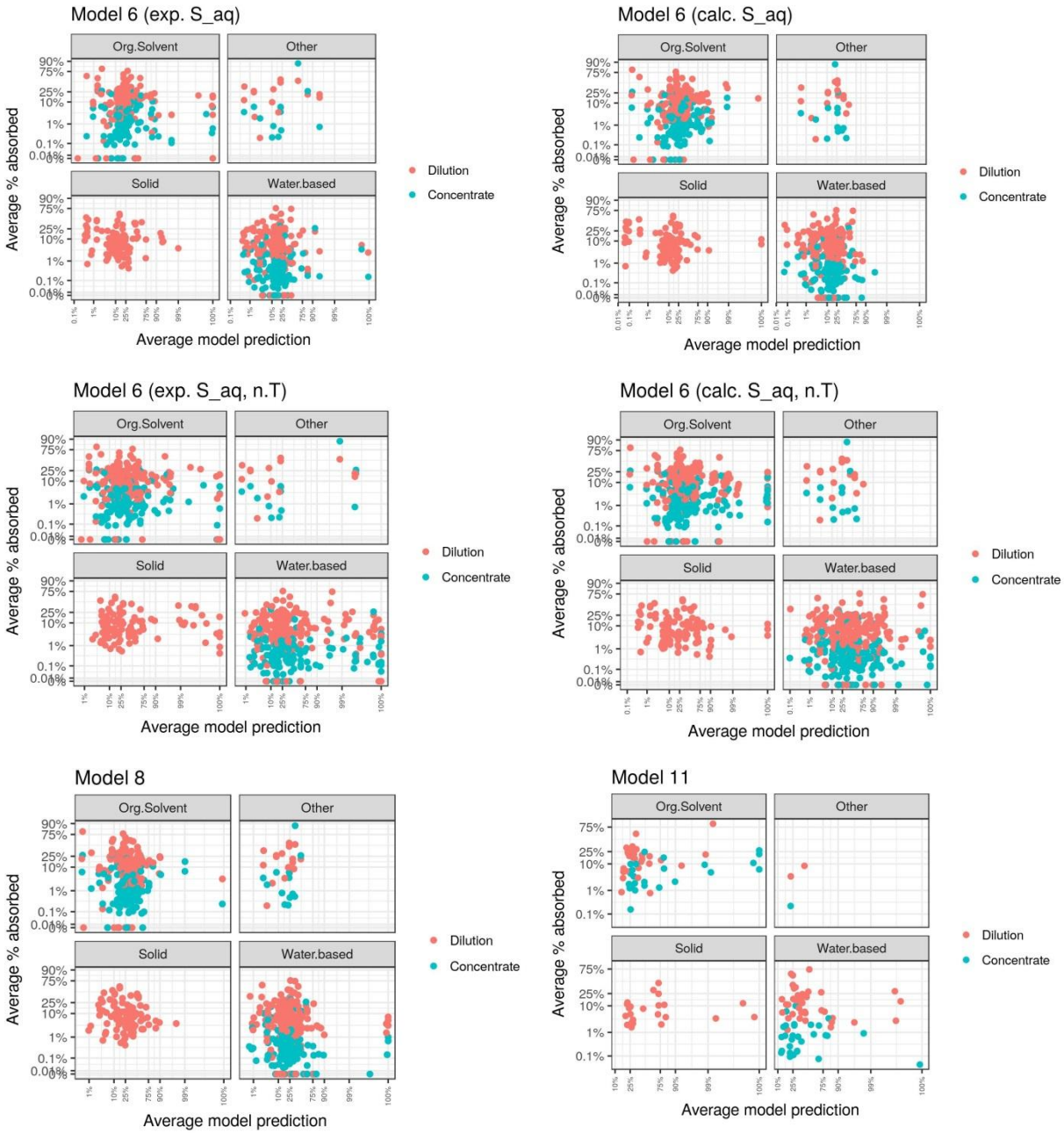


Figure L.2: Relationships between the averaged experimental % Dermal Absorption (%DA) values and the predicted ones divided by formulations groups and into concentrates and dilutions. For models' explanations please refer to Appendix G.

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Other covariates

Below is presented the covariate analysis performed for models 7, 9 and 10. The sub-ranges were built according to the methodology outlined in 2.8.2.2. Figure L.3 and Figure L.4 provide an overview of the rank correlations (Spearman and Kendall tau) inside each model subset.

Table L.1: Covariates, their 4 sub-ranges and counts of data points per sub-range for model 7.

covariate	**MW**			
sub-range	[112,230]	(230,301]	(301,377]	(377,732]
n	60	57	59	57
covariate	**LogPow**			
sub-range	[-4.36,0.67]	(0.67,2.89]	(2.89,3.7]	(3.7,8.1]
n	59	64	59	51
covariate	**Saq**			
sub-range	[0,6.8)	[6.8,34.4)	[34.4,780)	[780,1.44e+08]
n	56	58	57	62
covariate	**pKa**			
sub-range	[-7.8,1.95]	(1.95,3.13]	(3.13,9.4]	(9.4,17.9]
n	63	54	59	55
covariate	**Charge (QO.N [e])**			
sub-range	[0.3,0.64)	[0.64,0.83)	[0.83,1.2)	[1.2,4.56]
n	52	52	59	70
covariate	**HB**			
sub-range	[1,4)	[4,5)	[5,7)	[7,13]
n	54	53	55	71
covariate	**MP**			
sub-range	[-170,64.2)	[64.2,116)	[116,162)	[162,360]
n	55	61	54	63
covariate	**Reactivity**			
sub-range	[6.69,8.55]	(8.55,8.97]	(8.97,9.44]	(9.44,12.2]
n	61	57	62	53
covariate	**Concentration (mg/L)**			
sub-range	[0.0156,0.52]	(0.52,3.24]	(3.24,127]	(127,800]
n	59	58	58	58
covariate	**Concentrate**			
sub-range	FALSE		TRUE	
n	145		88	
covariate	**Merged Form**			
sub-range	Organic Solvent	Other	Solid	Water based
n	74	12	61	84

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

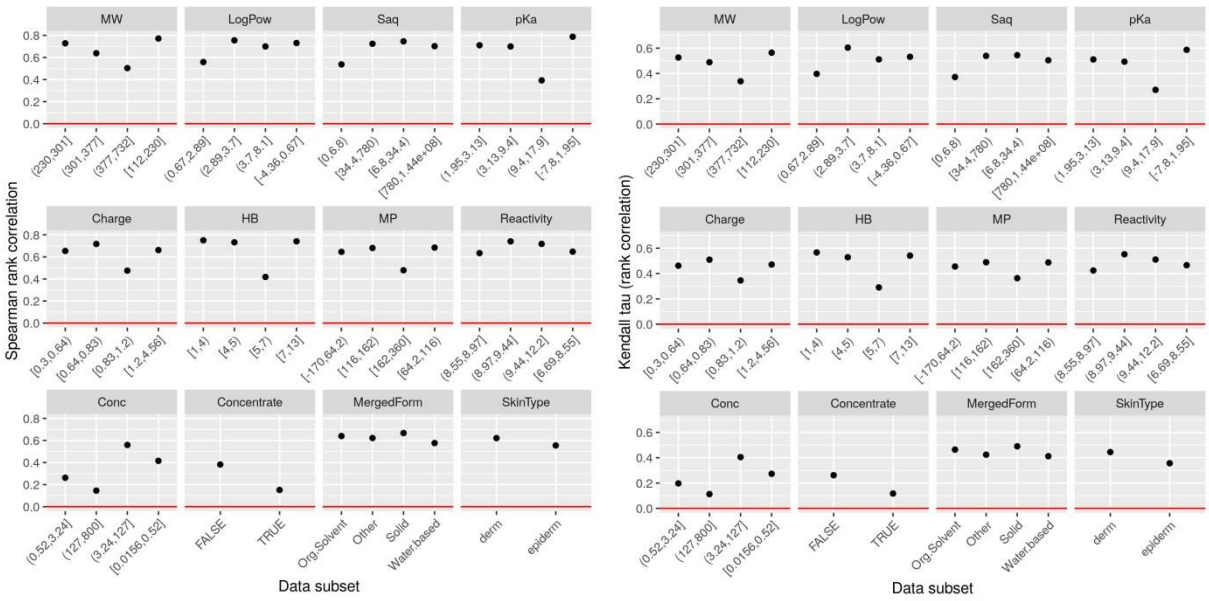


Figure L.3: Model 7: Spearman (left) and Kendall tau (right) rank correlations for the subsets presented in Table L.1

Table L.2: Covariates, their 4 sub-ranges and counts of data points per sub-range for Models 9 and 10.

covariate	**MW**			
sub-range	[112,249]	(249,330]	(330,401]	(401,1.63e+03]
n	41	37	46	31
covariate	**LogPow**			
sub-range	[-1.96,1.25]	(1.25,3.2]	(3.2,4.11]	(4.11,6.2]
n	40	45	31	39
covariate	**Saq**			
sub-range	[0,1.38]	(1.38,7.3]	(7.3,200]	(200,1.44e+08]
n	45	34	40	36
covariate	**pKa**			
sub-range	[-7.19,1.07]	(1.07,4.17]	(4.17,10.9]	(10.9,17.3]
n	39	38	40	34
covariate	**Charge (QO.N [e])**			
sub-range	[0,0.59]	(0.59,0.78]	(0.78,1.07]	(1.07,1.99]
n	40	38	41	36
covariate	**HB**			
sub-range	[2,4]	(4,5]	(5,7]	(7,16]
n	52	35	45	23
covariate	**MP**			
sub-range	[-61.1,76]	(76,118]	(118,157]	(157,298]
n	41	38	37	39
covariate	**Reactivity**			
sub-range	[7.12,8.31]	(8.31,8.66]	(8.66,8.97]	(8.97,12.2]
n	39	43	33	39
covariate	**Concentration (mg/l)**			

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

sub-range	[0.0156,0.58]	(0.58,2.5]	(2.5,111]	(111,712]
n	39	39	38	39
covariate	**Concentrate**			
sub-range	FALSE		TRUE	
n	99		56	
covariate	**Merged Form**			
sub-range	Organic Solvent	Other	Solid	Water based
n	55	3	23	72

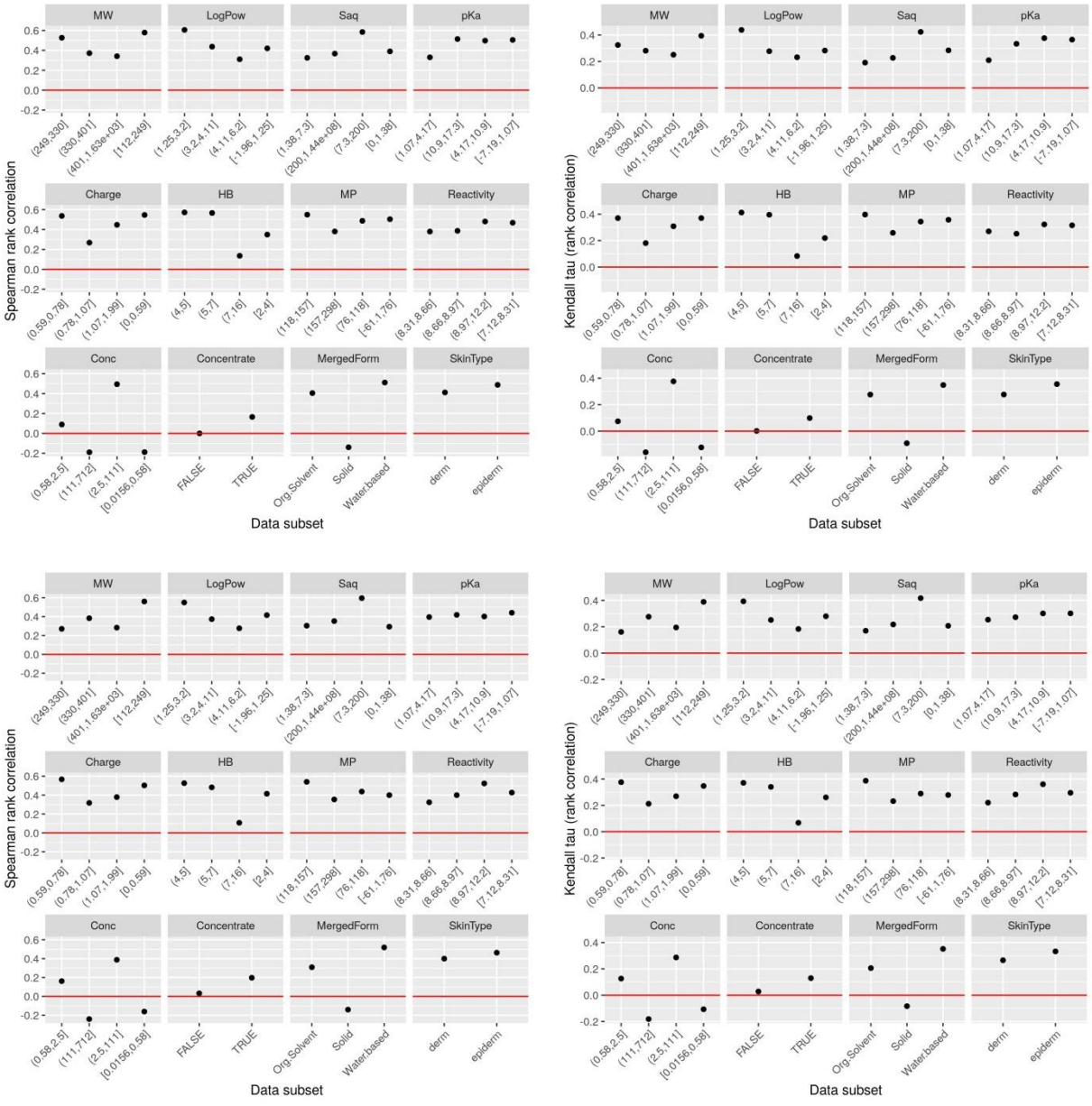


Figure L.4: Model 9 (top) and model 10 (bottom): Spearman (left) and Kendall tau (right) rank correlations for the subsets presented in Table L.2

The present document has been produced and adopted by the bodies identified above as authors. This task has been carried out exclusively by the authors in the context of a contract between the European Food Safety Authority and the authors, awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.